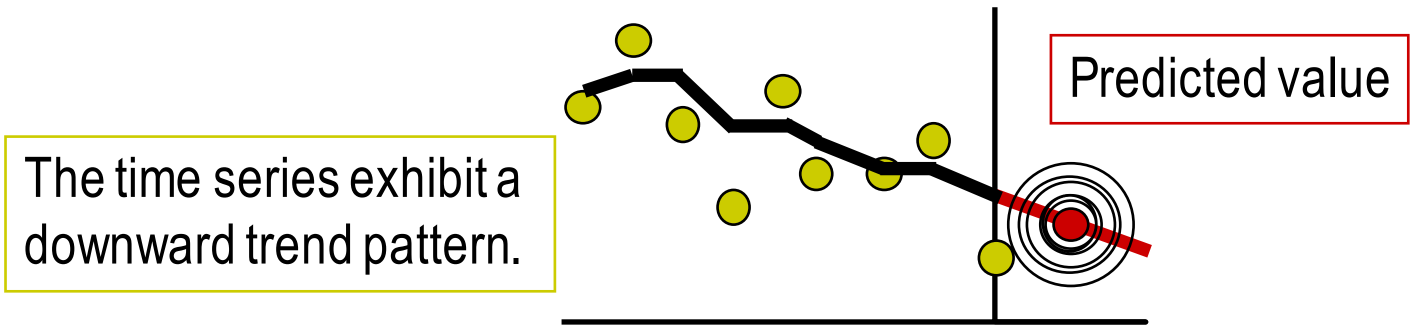
Ch 20 實習

Agenda

- 時間序列概念
- 四種模型算法
  - 移動平均法（Moving Averages）
  - 指數平滑法 （Exponentially Smoothed Time Series）
  - 趨勢與季節性分析法 （Trend or Seasonal Effects）
  - 自我相關分析（ Autoregressive model）
- 下一期預測
- 比較預測模型MAD與ＳＳＥ法

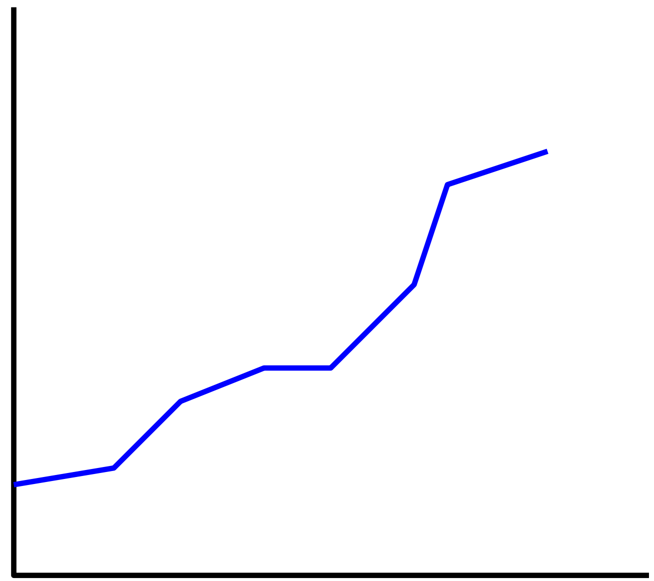
1. 時間序列概念

- Any variable that is measured over time in sequential order is called a **time series.**
- We analyze time series to detect patterns.
- The patterns help in forecasting future values of the time series.
- 回歸預測是線性的，但是資料本身可能因為TCSR而造成預測上的誤差，或是研究者難以判斷趨勢，所以要修正他，使下一期預測能更精準或是更容易判斷。



1.時間序列概念

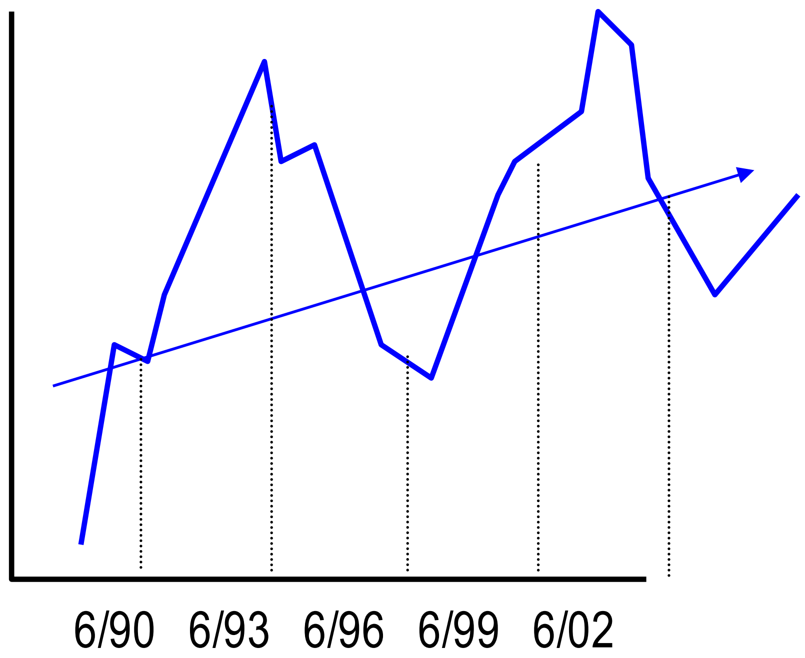
- A **time series** can consist of four components.
  - Long - term trend (T).
  - Cyclical effect (C).
  - Seasonal effect (S).
  - Random variation (R).



A trend is a long term relatively smooth pattern or direction, that persists usually for more than one year.

1.時間序列概念

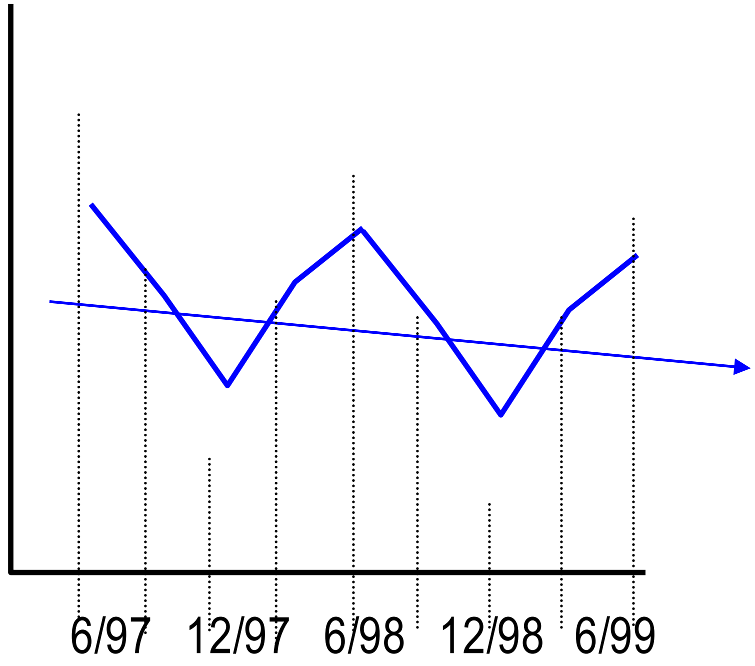
- A **time series** can consists of four components.
  - Long - term trend (T)
  - Cyclical variation (C)
  - Seasonal variation (S)
  - Random variation (R)



A **cycle** is a wavelike pattern describing a long term behavior (for more than one year). 是長期的趨勢～例如景氣循環
Cycles are seldom regular, and often appear in combination with other components

1.時間序列概念

- A **time series** can consists of four components.
  - Long - term trend (T).
  - Cyclical effect (C).
  - Seasonal effect (S).
  - Random variation (R).

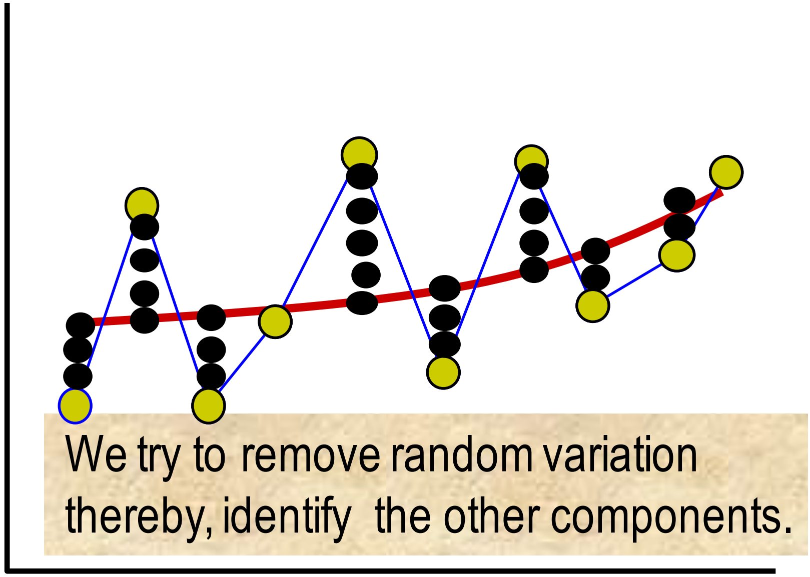


The **seasonal component** of the time series exhibits a short term (less than one year) calendar repetitive behavior. 是短期的趨勢～冷氣機的需求

1.時間序列概念

- A **time series** can consists of four components.
  - Long - term trend (T).
  - Cyclical effect (C).
  - Seasonal effect (S).
  - Random variation (R).

Random variation comprises the irregular unpredictable changes in the time series. It tends to hide the other (more predictable) components. 妨礙你看到趨勢的雜質



2.四種模型算法

- To identify the components present in the time series, we need first to remove the random variation.
- 因為要看出趨勢，可能受到雜質的影響，所以需要利用smoothing techniques來減低這些雜質的影響。
  - 移動平均法（Moving Averages）
  - 指數平滑法（Exponentially Smoothed Time Series）

2.1 移動平均法 （Moving Averages）

- A **k-period moving average** for time period t is the *arithmetic average* of the time series values around period t.
  - For example: A 3-period moving average at period t is calculated by $(y_{t-1} + y_t + y_{t+1})/3$

2.1 移動平均法 （Moving Averages）

- To forecast future gasoline sales, the last four years quarterly sales were recorded.
- Calculate the three-quarter and five-quarter moving average.

Period	Year/Quarter		Gas Sales
1	1	1	39
2		2	37
3		3	61
4		4	58
5	2	1	18
6		2	56

2.1 移動平均法 （Moving Averages）

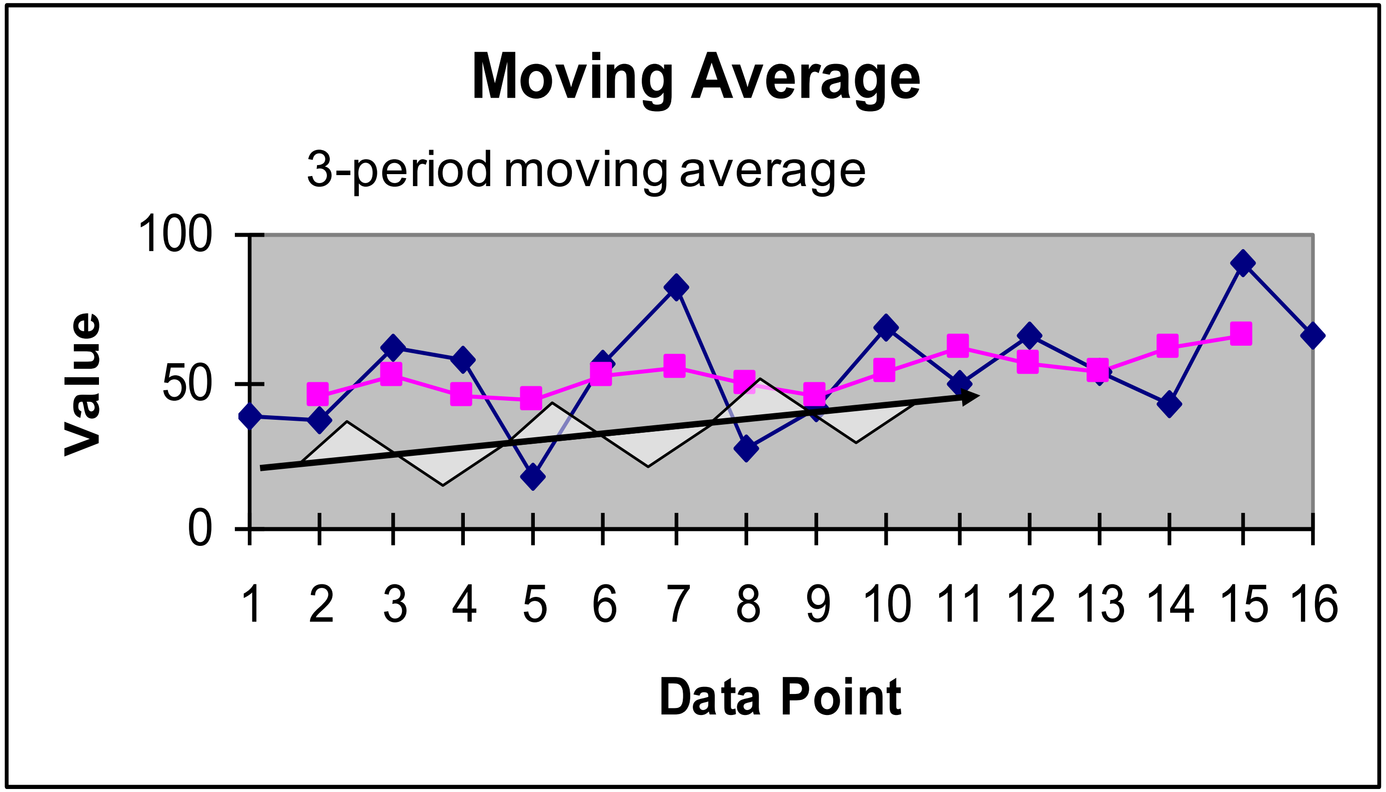
- Solution
  - Solving by hand

(39+37+61)/3=

(39+37+61+58+18)/5=

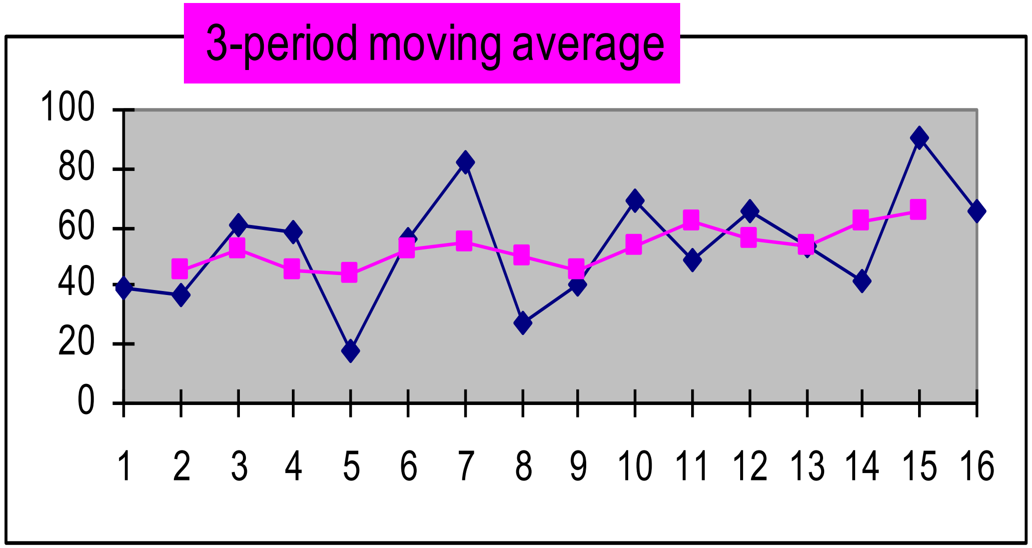
Period	Gas Sales	3-period moving Avg	5-period moving Avg
1	39	*	*
	37	45.6667	*
	61	52.0000	42.6000
4	58	45.6667	46.0000
5	18	44.0000	55.0000
6	56	52.0000	48.2000
7	82	55.0000	44.8000
8	27	50.0000	55.0000
9	41	45.6667	53.6000
10	69	53.0000	50.4000
11	49	61.3333	55.8000
12	66	56.3333	56.0000
13	54	54.0000	60.2000
14	42	62.0000	63.6000
15	90	66.0000	*
16	66	*	*

2.1 移動平均法 （Moving Averages）



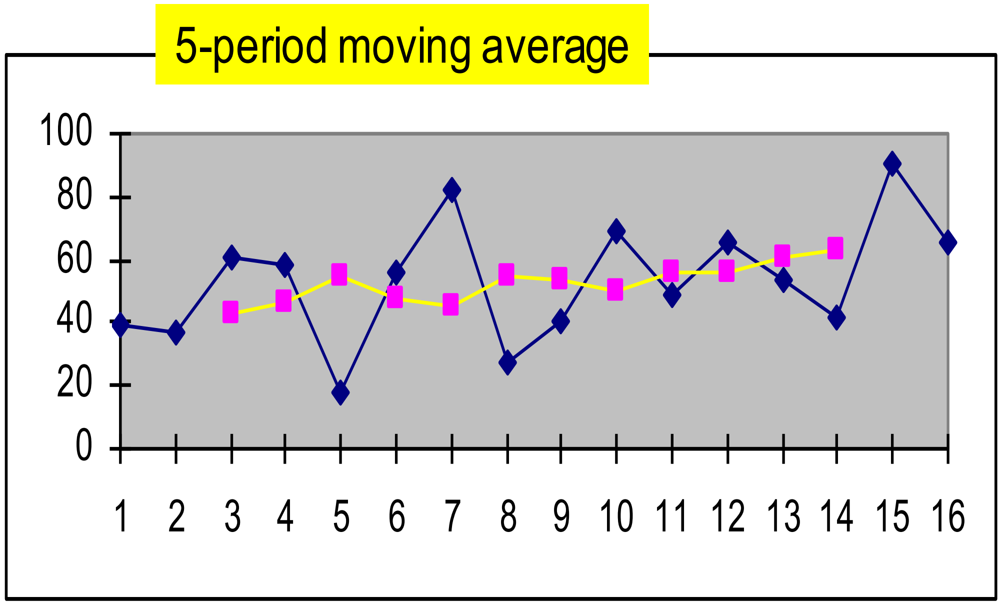
Notice how the averaging process removes some of the random variation.
Comment: We can identify a trend component and seasonality as well.

2.1 移動平均法 （Moving Averages）



The 5-period moving average removes more variation than the 3-period moving average.

Too much smoothing may eliminate patterns of interest.
Here, the seasonality is removed when using 5-period moving average.



Too little smoothing leaves much of the variation, which disguises the real patterns.

五期移動平均雖然可以刪除較多變異（比較容易看出趨勢），可是季節性的資訊也會不見了。

Example 1

- For the following time series, compute the three-period moving averages.

Period	Time Series	Period	Time Series
1	48	7	43
2	41	8	52
3	37	9	60
4	32	10	48
5	36	11	41
6	31	12	30

Solution 1

Time series	Moving average
48	X
41	(48+41+37)/3 = 42.00
37	(41+37+32)/3 = 36.67
32	(37+32+36)/3 = 35.00
36	(32+36+31)/3 = 33.00
31	(36+31+43)/3 = 36.67
43	(31+43+52)/3 = 42.00
52	(43+52+60)/3 = 51.67
60	(52+60+48)/3 = 53.33
48	(60+48+41)/3 = 49.67
41	(48+41+30)/3 = 39.67
30	X

移動平均的缺點：資料回遺失
三期會有兩筆資料要犧牲，
五期有四筆資料要犧牲

所以才會有指數平滑法

2.1 Centered Moving Average

- With an even number of observations included in the moving average, the average is placed between the two periods in the middle.
- To place the moving average in an actual time period, we need to center it.
- Two consecutive moving averages are centered by taking their average, and placing it in the middle between them.

2.1 Centered Moving Average: Example

- Calculate the 4-period moving average and center it, for the data given below:

Period	Time series	Moving Avg.	Centerd Mov.Avg.
1	15		
2	27		
(2.5)		19.0	
3	20		20.25
(3.5)		21.5	
4	14		19.50
(4.5)		17.5	
5	25		
6	11		

2.2 指數平滑法 （Exponentially Smoothed Time Series）

$$S_t = wy_t + (1-w)S_{t-1}$$

S_t = exponentially smoothed time series at time t.

y_t = time series at time t.

S_{t-1} = exponentially smoothed time series at time t-1.

w = smoothing constant, where 0 £ w £ 1.

2.2 指數平滑法 (Exponentially Smoothed Time Series)

Calculate the gasoline sale smoothed time series using exponential smoothing with w = .2, and w = .7

Formula	Period	Gas Sales	Calculation
Set $S_1 = y_1$	1	39	$S_1 = 39$
$S_2 = wy_2 + (1\text{-}w)S_1$	2	37	$S_2 = (.2)(37) + (1\text{-}.2)(39) = 38.6$
$S_3 = wy_3 + (1\text{-}w)S_2$	3	61	$S_3 = (.2)(61) + (1\text{-}.2)(38.6) = 43.1$
	4	58	
	5	18	
	6	56	
	.	.	
	.	.	

Example 2

- Apply exponential smoothing with w=0.1 to help detect the components of the following time series

Period	1	2	3	4	5	6	7	8	9	10
Time Series	38	43	42	45	46	48	50	49	46	45

Solution 2

$S_t = wy_t + (1\text{-}w)S_{t-1}$

Time series	Exponentially smoothed time series
38	38
43	.1(43) +. 9(38) = 38.50
42	.1(42) +. 9(38.50) = 38.85
45	.1(45) +. 9(38.85) = 39.47
46	.1(46) +. 9(39.47) = 40.12
48	.1(48) +. 9(40.12) = 40.91
50	.1(50) +. 9(40.91) = 41.82
49	.1(49) +. 9(41.82) = 42.53
46	.1(46) + .9(42.53) = 42.88
45	.1(45) + .9(42.88) = 43.09

2.3 趨勢與季節性分析法 （Trend or Seasonal Effects）

- The trend component of a time series can be linear or non-linear.
- It is easy to isolate the trend component using linear regression.
  - For linear trend use the model $y = \beta_0 + \beta_1 t + \varepsilon$.
  - For non-linear trend with one (major) change in slope use the quadratic model $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon$

2.3 趨勢與季節性分析法 （Trend or Seasonal Effects）

- Seasonal variation may occur within a year or within a shorter period (month, week) 通常是短期
- To measure the seasonal effects we construct seasonal indexes. 需計算季節性(/週/月)指數
- Seasonal indexes express the degree to which the seasons differ from the average time series value across all seasons. 季節性指數，可以用來看當季是高於或低於平均的值

2.3 趨勢與季節性分析法 （Trend or Seasonal Effects）

- Remove the effects of the seasonal and random variations by regression analysis

跑趨勢分析求$\hat{y}$

$$\hat{y}_t = b_0 + b_1 t$$

- For each time period compute the ratio

計算比例$\frac{y}{\hat{y}}$

$$y_t/\hat{y}_t$$

which removes most of the trend variation

- For each season calculate the average of $y_t/\hat{y}_t$ which provides the measure of seasonality.

求每季的平均
（季節指數）

- Adjust the average above so that te sum of averages of all seasons is 1 (if necessary)

看季節指數是否需調整

2.3 Computing Seasonal Indexes

- Calculate the quarterly seasonal indexes for hotel **occupancy rate** in order to measure seasonal variation.

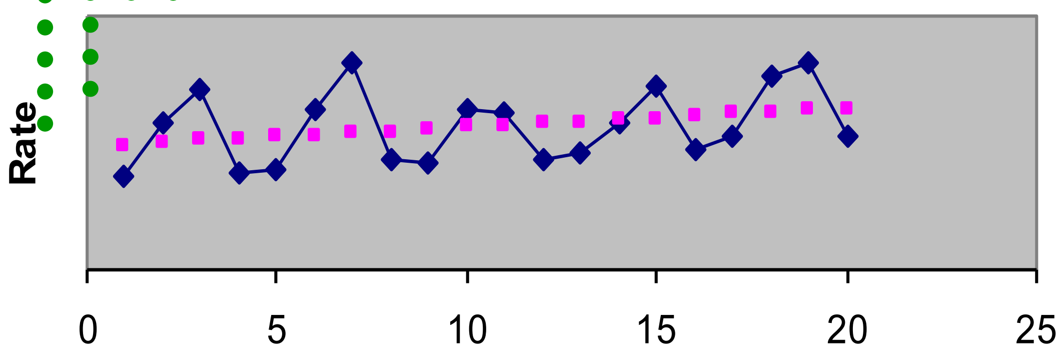
Year	Quarter	Rate	Year	Quarter	Rate	Year	Quarter	Rate
1996	1	0.561	1998	1	0.594	2000	1	0.665
	2	0.702		2	0.738		2	0.835
	3	0.8		3	0.729		3	0.873
	4	0.568		4	0.6		4	0.67
1997	1	0.575	1999	1	0.622			
	2	0.738		2	0.708			
	3	0.868		3	0.806			
	4	0.605		4	0.632			

2.3 Computing Seasonal Indexes

- Step1: Perform regression analysis for the model $y = \beta_0 + \beta_1 t + \varepsilon$ where t represents the time, and y represents the occupancy rate.

$$y=0.639368+0.005246t$$

Time (t)	Rate
1	0.561
2	0.702
3	0.800
4	0.568
5	0.575
6	0.738
7	0.868
8	0.605
.	.
.	.

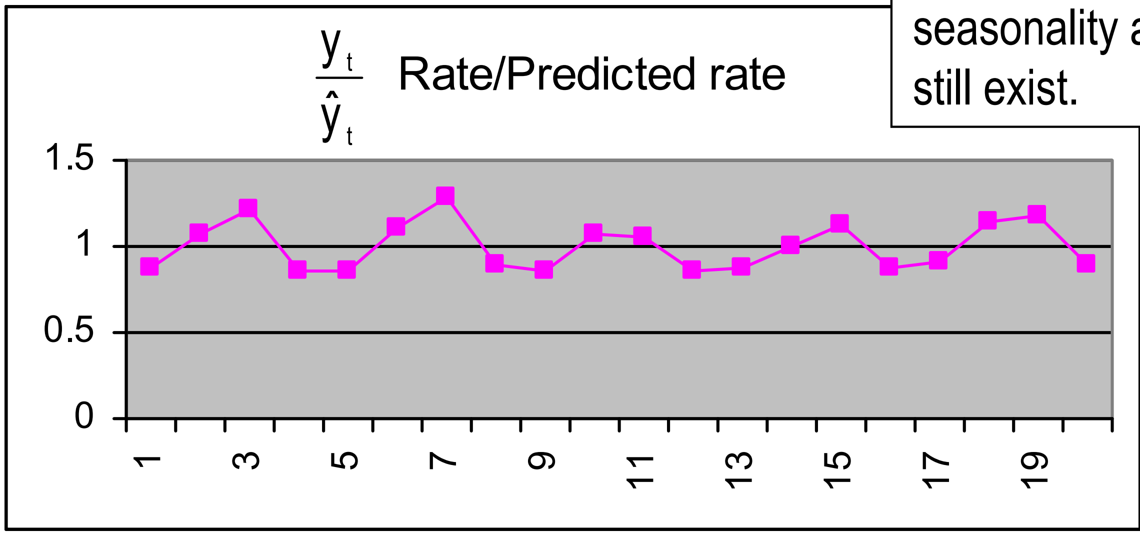


The regression line represents trend.

Step2: The Ratios

t	y_t	$\hat{y}_t$	Ratio
1	.561	.645	.561/.645=.870
2	.702	.650	.702/.650=1.08
3	....	....	....

=.639368+.005245(1)

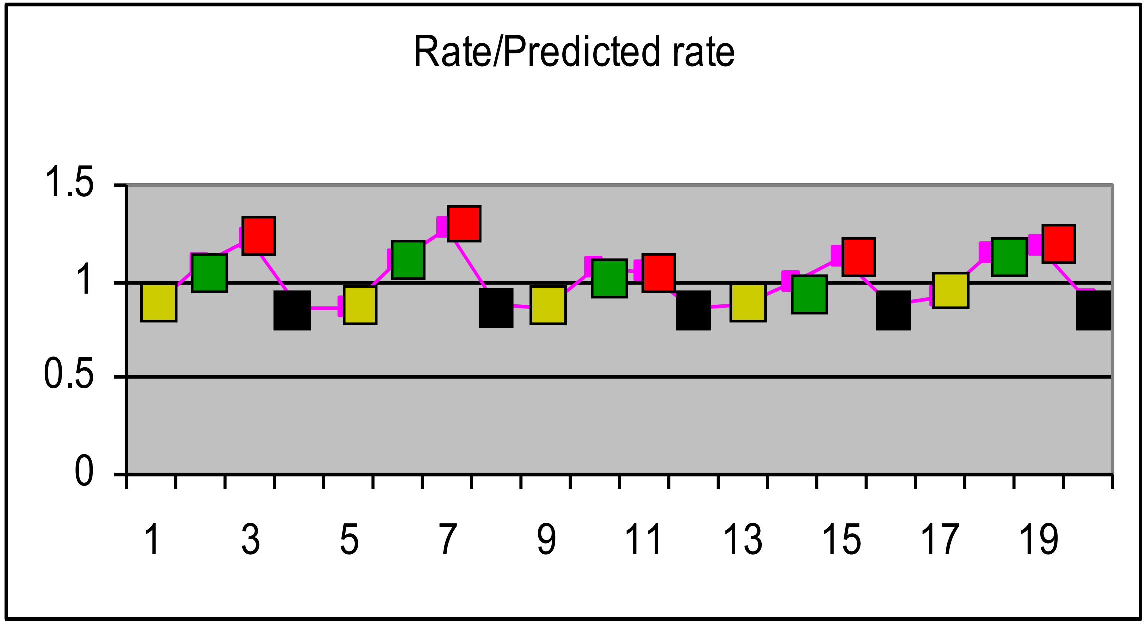


No trend is observed, but seasonality and randomness still exist.

Step3: The Average Ratios by Seasons

Rate/Predicted rate
0.870
1.080
1.221
0.860
0.864
1.100
1.284
0.888
0.865
1.067
1.046
0.854
0.879
0.993
1.122
0.874
0.913
1.138
1.181
0.900

- To remove most of the random variation but leave the seasonal effects, average the terms $y_t / \hat{y}_t$ for each season.



Average ratio for quarter 1: (.870 + .864 + .865 + .879 + .913)/5 = .878

Average ratio for quarter 2: (1.080+1.100+1.067+.993+1.138)/5 = 1.076

Average ratio for quarter 3: (1.221+1.284+1.046+1.122+1.181)/5 = 1.171

Average ratio for quarter 4: (.860 +.888 + .854 + .874 + .900)/ 5 = .875

Step4: Adjusting the Average Ratios

- In this example the sum of all the averaged ratios must be 4, such that the average ratio per season is equal to 1.
- If the sum of all the ratios is not 4, we need to adjust them proportionately.

In our problem the sum of all the averaged ratios is equal to 4:
.878 + 1.076 + 1.171 + .875 = 4.0.
No normalization is needed. These ratios become the seasonal indexes.

Suppose the sum of ratios is equal to 4.1. Then each ratio will be multiplied by 4/4.1

$$\text{Seasonal index} = \frac{(\text{Seasonal averaged ratio})\,(\text{number of seasons})}{\text{Sum of averaged ratios}}$$

Example 3

- For the following time series, compute the seasonal (daily) index. The regression line is

$\hat{y} = 16.8 + 0.366t$ (t=1,2,…,20)

	Week			
Day	1	2	3	4
Monday	12	11	14	17
Tuesday	18	17	16	21
Wednesday	16	19	16	20
Thursday	25	24	28	24
Friday	31	27	25	32

Solution 3 (1)

Step1:跑趨勢分析可得Yhat

$$\hat{y} = b0 + b1t + \varepsilon$$

$$\hat{y} = 16.8 + 0.366t + \varepsilon$$

Step 2: 算ratio

Week	Day	Period t	y	$\hat{y}$	$y/\hat{y}$
1	1	1	12	17.2	0.699
	2	2	18	17.5	1.027
	3	3	16	17.9	0.894
	4	4	25	18.3	1.369
	5	5	31	18.6	1.664
2	1	6	11	19.0	0.579
	2	7	17	19.4	0.878
	3	8	19	19.7	0.963
	4	9	24	20.1	1.194
	5	10	27	20.5	1.320
3	1	11	14	20.8	0.672
	2	12	16	21.2	0.755
	3	13	16	21.6	0.742
	4	14	28	21.9	1.277
	5	15	25	22.3	1.122
4	1	16	17	22.7	0.750
	2	17	21	23.0	0.912
	3	18	20	23.4	0.855
	4	19	24	23.8	1.010
	5	20	32	24.1	1.327

Solution 3 (2)

Step 3: 算週指數（把星期一取平均，星期二取平均…)

	Day					
Week	Monday	Tuesday	Wednesday	Thursday	Friday	Total
1	.699	1.027	.894	1.369	1.664	
2	.579	.878	.963	1.194	1.320	
3	.672	.755	.742	1.277	1.122	
4	.750	.912	.855	1.010	1.327	
Average	.675	.893	.864	1.213	1.358	5.003
Seasonal Index	.675	.892	.864	1.212	1.357	5.000

Step 4: 調整指數（因為相加不為五，所以要調整）

$\frac{0.675*5}{5.003}$=0.6745 (取0.675)

$\frac{1.358*5}{5.003}$=1.3571 (取1.357)

Solution 3 (3)

	Day					
Week	Monday	Tuesday	Wednesday	Thursday	Friday	Total
1	.699	1.027	.894	1.369	1.664	
2	.579	.878	.963	1.194	1.320	
3	.672	.755	.742	1.277	1.122	
4	.750	.912	.855	1.010	1.327	
Average	.675	.893	.864	1.213	1.358	5.003
Seasonal Index	.675	.892	.864	1.212	1.357	5.000

Step 4: 表示星期一到三的銷售是低於平均，四與五是高於平均

3.下一期預測（Forecasting）

- The choice of a forecasting technique depends on the components identified in the time series.
- The techniques discussed next are:
  - Exponential smoothing
  - Seasonal indexes
  - Autoregressive models (a brief discussion)

3.1 Forecasting with Exponential Smoothing

- The exponential smoothing model can be used to produce forecasts when the time series...
  - exhibits gradual(not a sharp) trend
  - no cyclical effects
  - no seasonal effects
- Forecast for period t+k is computed by

$$F_{t+k} = S_t$$ 後面預測都是一樣的值

where t is the current period;

$$S_t = \omega y_t + (1-\omega)S_{t-1}$$

Example 2

$S_t = wy_t + (1-w)S_{t-1}$

Time series	Exponentially smoothed time series
38	38
43	.1(43) +. 9(38) = 38.50
42	.1(42) +. 9(38.50) = 38.85
45	.1(45) +. 9(38.85) = 39.47
46	.1(46) +. 9(39.47) = 40.12
48	.1(48) +. 9(40.12) = 40.91
50	.1(50) +. 9(40.91) = 41.82
49	.1(49) +. 9(41.82) = 42.53
46	.1(46) + .9(42.53) = 42.88
45	.1(45) + .9(42.88) = 43.09

S10=43.09

預測 S11=43.09

預測 S12=43.09

3.2 Forecasting with Seasonal Indexes

- Linear regression and seasonal indexes to forecast the time series that composes trend with seasonality
- The model

$$F_t = [b_0 + b_1 t] \cdot SI_t$$

Linear trend value for period t, obtained from the linear regression

Seasonal index for period t.

3.3 Computing Seasonal Indexes

- Calculate the quarterly seasonal indexes for hotel **occupancy rate** in order to measure seasonal variation. 如何預測2001年的四筆資料(y21-y24)?

Year	Quarter	Rate	Year	Quarter	Rate	Year	Quarter	Rate
1996	1	0.561	1998	1	0.594	2000	1	0.665
	2	0.702		2	0.738		2	0.835
	3	0.8		3	0.729		3	0.873
	4	0.568		4	0.6		4	0.67
1997	1	0.575	1999	1	0.622			
	2	0.738		2	0.708			
	3	0.868		3	0.806			
	4	0.605		4	0.632			

$$\hat{y}_t = .639 + .00525t$$

季節指數：

S1:0.878

S2: 1.076

S3: 1.171

S4=0.875

為什麼是預測y21-y24?

3.3 Forecasting with Seasonal Indexes: Example

- The trend line was obtained from the regression analysis.

$$\hat{y}_t = .639 + .00525t$$

$$F_t = [b_0 + b_1 t] \cdot SI_t$$

– For the year 2001 we have:

$$\hat{y}_{21} = .639 + .00525\ (21) = .749$$

$$F_{21} = .749(.878)$$

t	Trend value	Quarter	SI	Forecast
21	.749	1	.878	**.658**
22	.755	2	1.076	**.812**
23	.760	3	1.171	**.890**
24	.765	4	.875	**.670**

Example 4

- The following trend line and seasonal indexes were computed from 10 years of quarterly observations. Forecast the next year's time series.

$$\hat{y} = 150 + 3t \qquad t = 1, 2, \ldots, 40$$

Quarter	Seasonal Index
1	0.7
2	1.2
3	1.5
4	0.6

所以要預測哪幾期？

Solution 4

Y41=150+3(41)=273

Y41=273*0.7=191.1

Quarter	t	$\hat{y} = 150 + 3t$	SI	Forecast
1	41	273	.7	191.1
2	42	276	1.2	331.2
3	43	279	1.5	418.5
4	44	282	.6	169.2

$$F_t = [b_0 + b_1 t] \cdot SI_t$$

3.3 Autoregressive models

- Autocorrelation among the errors of the regression model provides opportunity to produce accurate forecasts.
- In a stationary time series (no trend and no seasonality) correlation between consecutive residuals leads to the following autoregressive model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

Autoregressive models

- Example 20.6 (Xm20-06)
  - Forecast the increase in the Consumer Price Index (CPI) for the year 2000, based on the annual percentage increase in the CPI collected for the years 1980 - 1990.

t	y(t)
1	13.5
2	10.4
3	6.1
4	3.2
.	.
.	.

Autoregressive model; Example

- The estimated model has the form

$$y_t = b_0 + b_1 y_{t-1}$$

Y (t)	Y (t-1)
13.5	
10.4	13.5
6.1	10.4
3.2	6.1
4.3	3.2
.	4.3
.	.

The increase in the CPI for periods 1, 2, 3,… are predictors of the increase in the CPI for periods 2, 3, 4,…, respectively.

Autoregressive model; Example

為何是t=21? 因為1980-2000期間，2000年剛好是第21筆資料

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8748
R Square	0.7653
Adjusted R Square	0.7515
Standard Error	0.9898
Observations	19

$$\hat{y}_t = 1.20 + .59 y_{t-1}$$

$$Forecast \ \ for \ 2000 \ (t = 21):$$

$$\hat{y}_{21} = 1.20 + .595 y_{20}$$

$$= 1.20 + .59(2.2) = 2.50$$

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	54.31	54.31	55.44	0.0000
Residual	17	16.66	0.98		
Total	18	70.97			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.20	0.42	2.90	0.0100
X Variable 1	0.59	0.08	7.45	0.0000

4. 比較預測模型

- A forecasting method can be selected by evaluating its forecast accuracy using the actual time series.

- The two most commonly used measures of forecast accuracy are:
  - Mean Absolute Deviation

$$MAD = \frac{\sum_{t=1}^{n} |y_t - F_t|}{n}$$

  - Sum of Squares for Forecast Error

$$SSE = \sum_{t=1}^{n} (y_t - F_t)^2$$

Example 5

5. Two forecasting models were used to predict the future values of a time series. These are shown in the accompanying table, together with the actual values.

Forecast Value F_t Model 1	Model 2	*Actual Value y_t*
9.0	7.7	7.6
7.8	8.1	8.2
7.0	8.5	8.9
9.6	9.0	11.0

Compute the mean absolute deviation (MAD) and sum of squares for forecast (SSE) for each model to determine which was more accurate.

Solution 5

$$MAD = \frac{\sum_{t=1}^{n} |y_t - F_t|}{n}$$

$$MAD1 = \frac{|7.6-9|+|8.2-7.8|+|8.9-7.0|+|11-9.6|}{4} = 1.275$$

$$SSE = \sum_{t=1}^{n} (y_t - F_t)^2$$

$$SSE1 = (7.6-9)^2 + (8.2-7.8)^2 + (8.7-7)^2 + (11-9.6)^2 = 7.69$$

5.

Model	MAD	SSE
1	1.275	7.69
2	0.650	4.18

Model 2 was more accurate since it had the smallest MAD and SSE values.