Day 9

Forecasting

Nonlinear Trend Regression Example
콜레스테롤약 판매 비선형추세 예제

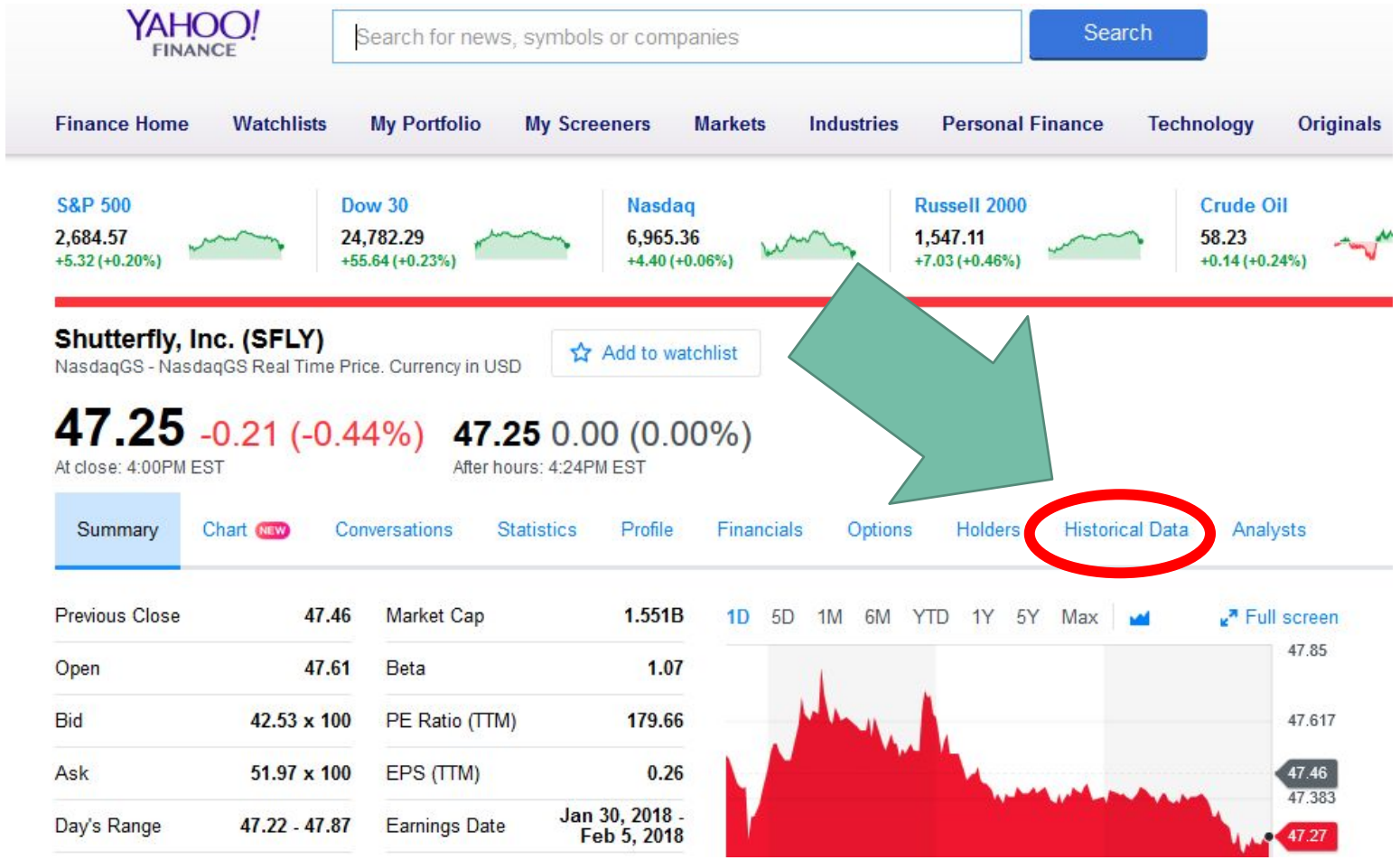
- This table shows the revenues for a cholesterol drug since the company won FDA approval for it 10 years ago. Forecast the revenues of cholesterol drug for year 11 and 12.
  테이블은 지난 10년간 콜레스테롤약 판매를 보여줍니다. 11년, 12년의 판매수익을 예측하시요

Year	Revenue
1	23.1
2	21.3
3	27.4
4	34.6
5	33.8
6	43.2
7	59.5
8	64.4
9	74.2
10	99.3
11	?
12	?

Exercise #9

- Import the revenue data **콜레스테롤약 판매수익 데이터를 가져옴**
- Plot the time series **시계열차트**
- Conduct a regression analysis and plot the regression equation on the time series plot **회귀분석, 회귀분석차트**
- Forecast year 11 and year 12 **11년, 12년 예측**
- Create an error table and calculate ME, MAE, MAPE, and MSE **에러테이블, 에러측정치 계산**
- Create a tracking signal table and plot the signals **추적신호와 그래프**

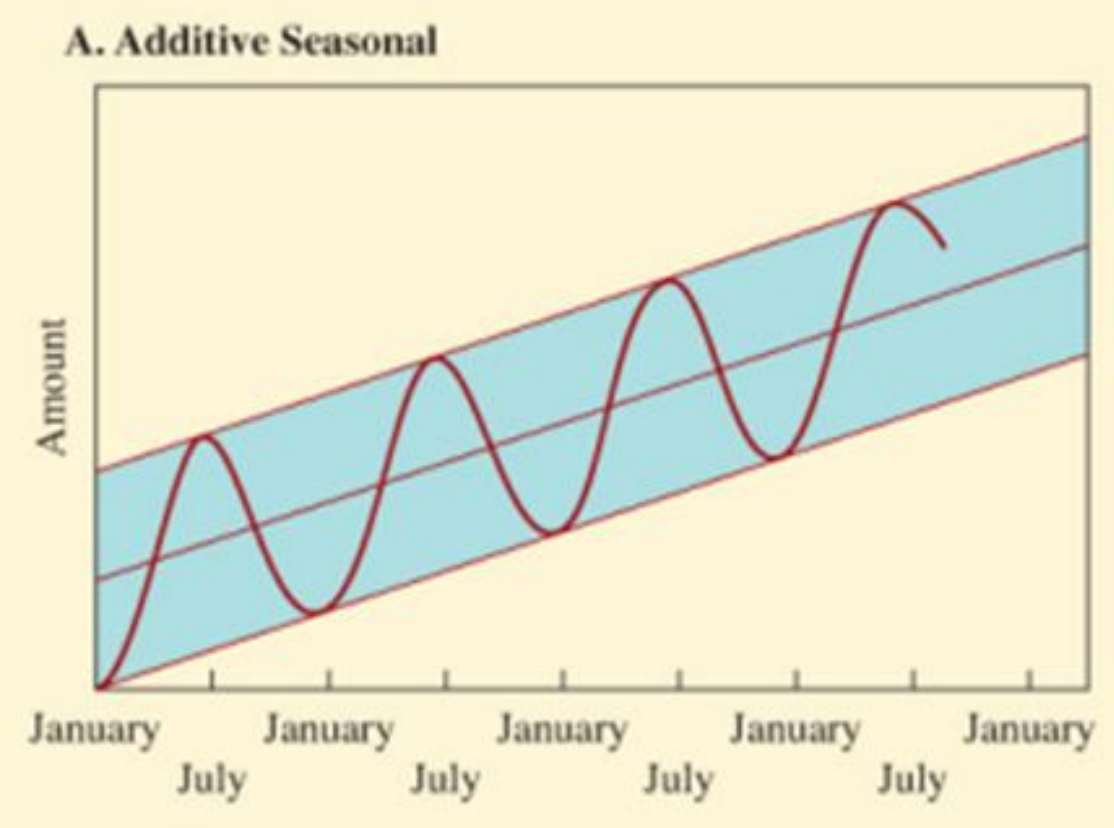
Stock Price Forecasting
주식가격 예측예제



Exercise #8

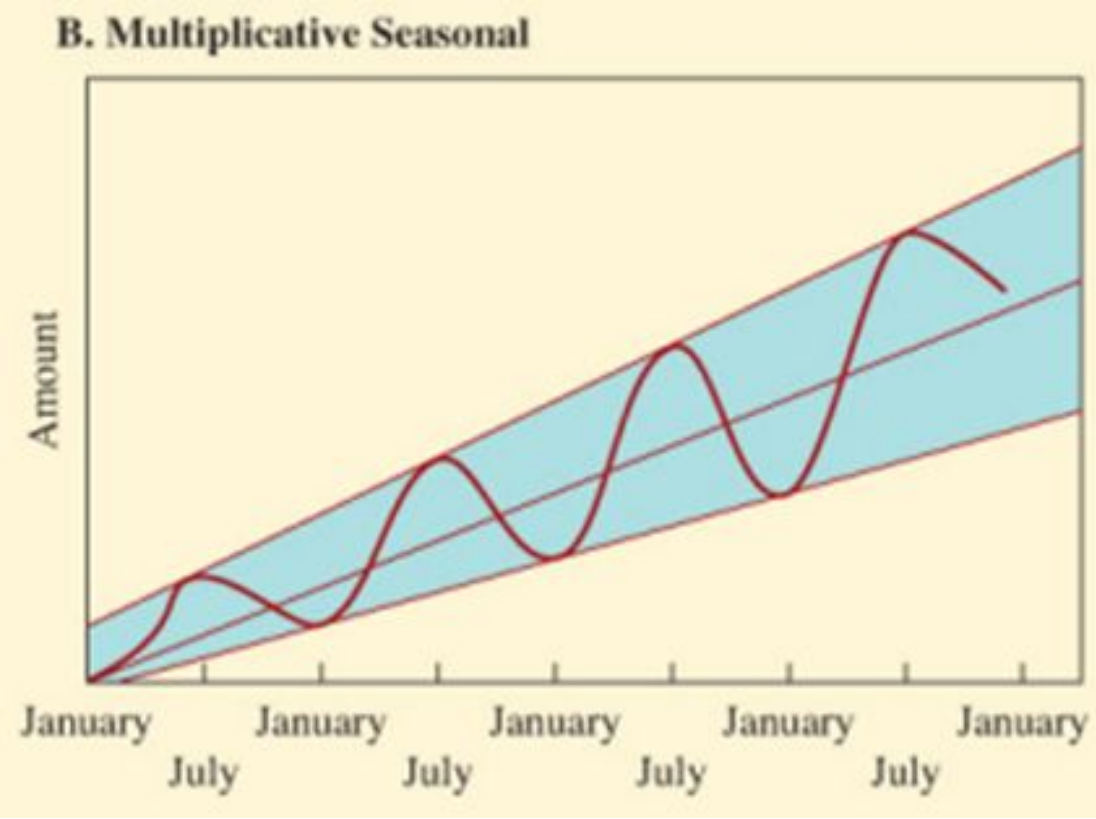
- Download the historical daily stock prices of Shutterfly from yahoo finance (approximately two years). Use the close prices for the analysis **셔터플라이의 일별 주식가격을 다운로드하여 종가로 분석**
- Plot the time series **시계열차트**
- Conduct a regression analysis and plot the regression equation on the time series plot **회귀분석, 회귀분석차트**
- Forecast next 30 days' stock price **다음 30일간의 주식가격 예측**
- Create an error table and calculate ME, MAE, MAPE, and MSE **에러테이블, 에러측정치 계산**
- Create a tracking signal table and plot the signals **추적신호와 그래프**

Additive Seasonal Forecast



- Additive seasonal variation assumes that seasonal amount is a constant no matter what the trend or average amount is.
  가법계절예측에서는 추세나 평균값과는 상관없이 계절요소가 같다고 봄
- Additive seasonal forecasting = trend + seasonal amounts

Multiplicative Seasonal Forecast



- In Multiplicative seasonal variation, the trend is multiplied by the seasonal factors.
  승법계절예측에서는 계절요소가 추세에 곱해짐
- Multiplicative seasonal forecasting = trend x seasonal factor

Seasonal Forecast

- Simple Forecast with Seasonality 평균을 이용한 단순예측법
- Multiple Regression with Seasonality 더미변수를 이용한 다중회귀분석법
- Decomposition 분해법

Time Series Components
시계열요소들

- A seasonal time series consists of a trend component, a seasonal component and an irregular component.
  계절시계열데이터는 추세, 계절, 불규칙요소로 구성됨

Trend 추세
- The gradual upward or downward movement of the data over time

Cycles 주기
- Patterns in the data that occur every several years

Seasonality 계절
- A data pattern that repeats itself after a period of days, weeks, months, or quarters

Random Variations 무작위변동
- "blips" in the data caused by chance and unusually situations

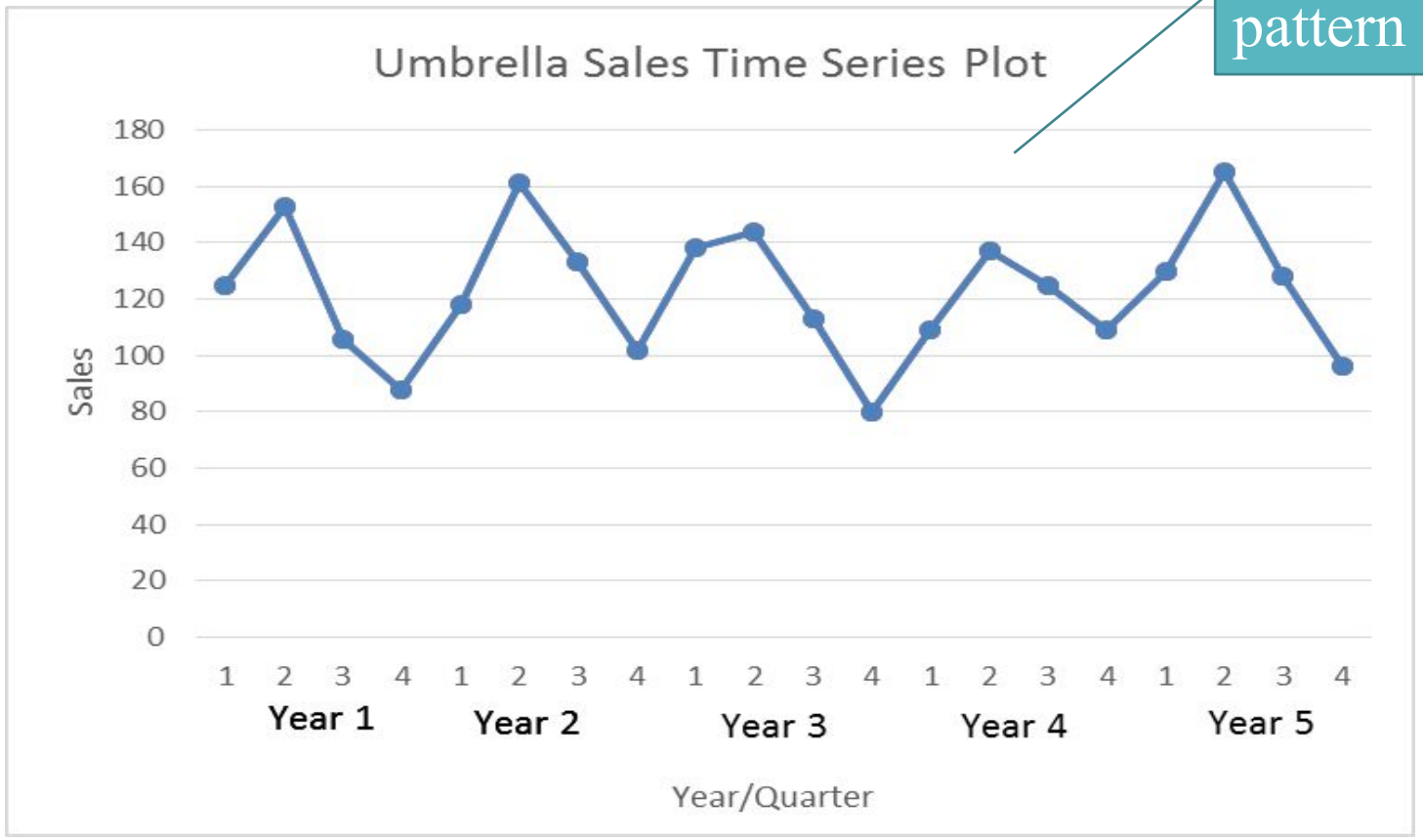
Umbrella Sales Example
우산세일 계절성 예측 예제

Year	Quarter	Sales
1	1	125
	2	153
	3	106
	4	88
2	1	118
	2	161
	3	133
	4	102
3	1	138
	2	144
	3	113
	4	80
4	1	109
	2	137
	3	125
	4	109
5	1	130
	2	165
	3	128
	4	96

This table contains the number of umbrella sold at a clothing store over the past five years. Forecast the quarterly sales for 6th year.

지난 5년간 옷가게에서 팔린 우산의 갯수가 테이블에 정리되어 있습니다. 내년 (year 6)의 쿼터별 세일을 예측해 보십시요.

Time Series Plot 시계열 라인차트



A quarterly seasonal pattern is present!

Simple Forecast with Seasonality
단순예측법

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
1	125	153	106	88
2	118	161	133	102
3	138	144	113	80
4	109	137	125	109
5	130	165	128	96
Average	124	152	121	95

You can obtain the quarterly forecasts for next year simply by computing the average number of umbrellas sold in each quarter! 단순하게 같은 기간을 평균냄

Multiple Regression with Seasonality
다중회귀분석법

- Can use a multiple regression to forecast the quarterly sales for next year. 다중회귀분석을 이용해서 예측

Estimated regression equation

+

Multiple Regression with Seasonality
가변수의 사용

Year	Quarter	Sales	Qrt1	Qrt2	Qrt3
1	1	125	1	0	0
	2	153	0	1	0
	3	106	0	0	1
	4	88	0	0	0
2	1	118	1	0	0
	2	161	0	1	0
	3	133	0	0	1
	4	102	0	0	0
3	1	138	1	0	0
	2	144	0	1	0
	3	113	0	0	1
	4	80	0	0	0
4	1	109	1	0	0
	2	137	0	1	0
	3	125	0	0	1
	4	109	0	0	0
5	1	130	1	0	0
	2	165	0	1	0
	3	128	0	0	1
	4	96	0	0	0
6	1		1	0	0
	2		0	1	0
	3		0	0	1
	4		0	0	0

- Treat the season as a categorical variable
  계절이나 쿼터를 범주형으로 봄
- When a categorical variable has k levels, k-1 dummy variables are required.
  범주보다 하나적게 더미변수를 설정

Estimated Regression Equation
추정회귀식

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.893791345							
R Square	0.798862968							
Adjusted R Square	0.761149775							
Standard Error	11.32475165							
Observations	20							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	8150	2716.666667	21.18258609	8.10363E-06			
Residual	16	2052	128.25					
Total	19	10202						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	95	5.064582905	18.75771446	2.5659E-12	84.26356386	105.7364361	84.26356386	105.7364361
Q1	29	7.162401832	4.048921113	0.000931211	13.8163864	44.1836136	13.8163864	44.1836136
Q2	57	7.162401832	7.958224258	5.93482E-07	41.8163864	72.1836136	41.8163864	72.1836136
Q3	26	7.162401832	3.630067205	0.002251556	10.8163864	41.1836136	10.8163864	41.1836136

Estimated Regression Equation

$$Sales = 95.0 + 29.0\,Qtr1 + 57.0\,Qtr2 + 26.0\,Qtr3$$

Forecasting Sales for 6th Year 예측값

	A	B	C	D	E	F	G	H	I
1	Year	Quarter	Sales	Qrt1	Qrt2	Qrt3	Forecast		
22	6	1		1	0	0	124		
23		2		0	1	0	152		
24		3		0	0	1	121		
25		4		0	0	0	95		
26									
27	SUMMARY OUTPUT								
28									
29	*Regression Statistics*								
30	Multiple R	0.89379134							
31	R Square	0.79886297							
32	Adjusted R Square	0.76114977							
33	Standard Error	11.3247517							
34	Observations	20							
35									
36	ANOVA								
37		*df*	*SS*	*MS*	*F*	*Significance F*			
38	Regression	3	8150	2716.66667	21.1825861	8.10363E-06			
39	Residual	16	2052	128.25					
40	Total	19	10202						
41									
42		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
43	Intercept	95	5.064582905	18.7577145	2.5659E-12	84.26356386	105.736436	84.2635639	105.736436
44	Qrt1	29	7.162401832	4.04892111	0.00093121	13.8163864	44.1836136	13.8163864	44.1836136
45	Qrt2	57	7.162401832	7.95822426	5.9348E-07	41.8163864	72.1836136	41.8163864	72.1836136
46	Qrt3	26	7.162401832	3.63006721	0.00225156	10.8163864	41.1836136	10.8163864	41.1836136

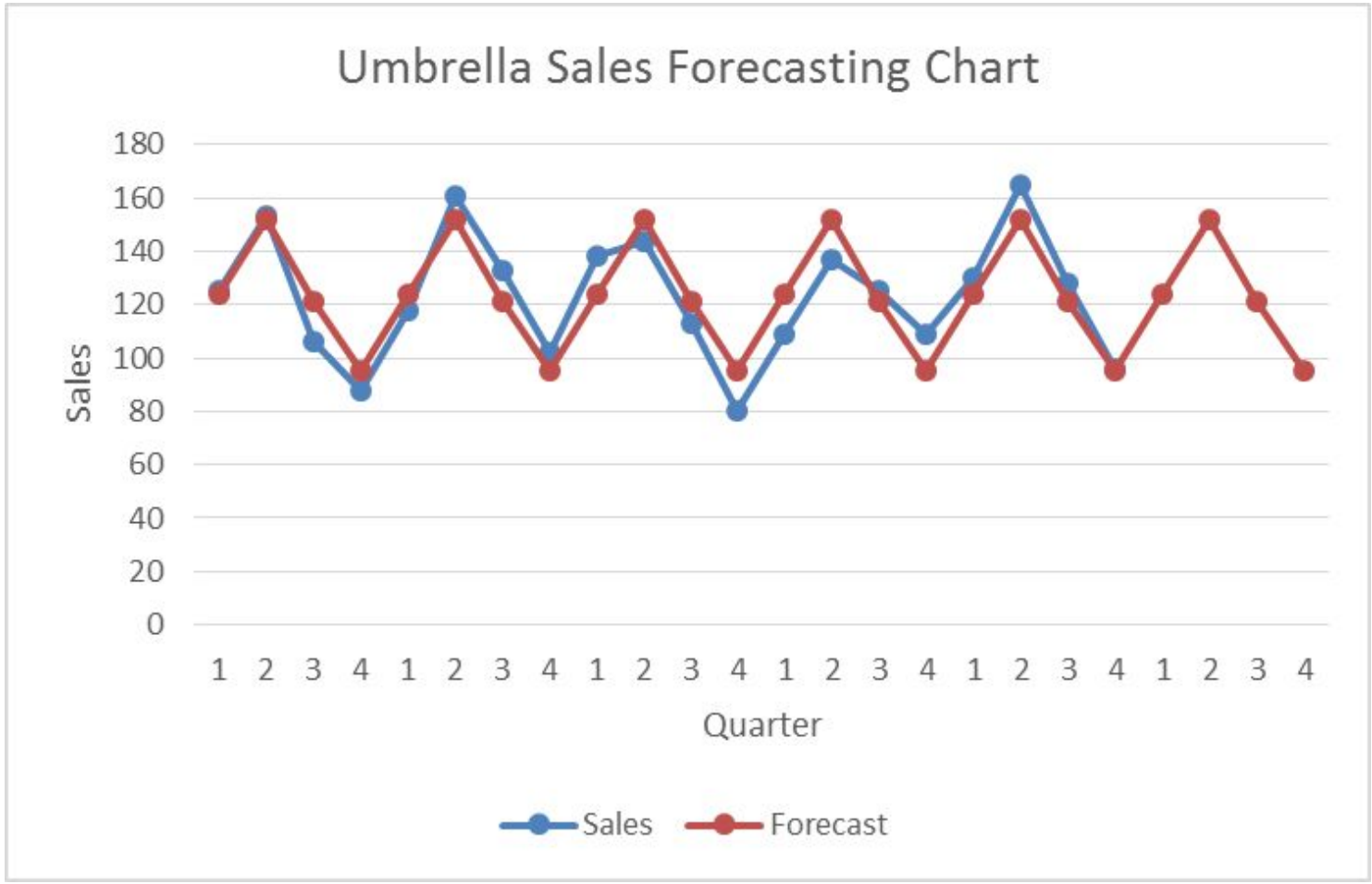
$$= 95.0 + 29.0 * 1 + 57.0 * 0 + 26.0 * 0$$

$$= 95.0 + 29.0 * 0 + 57.0 * 1 + 26.0 * 0$$

$$= 95.0 + 29.0 * 0 + 57.0 * 0 + 26.0 * 1$$

$$= 95.0 + 29.0 * 0 + 57.0 * 0 + 26.0 * 0$$

Forecast Chart 예측차트



Forecast Errors 예측오차

		ME	MAE	MAPE	MSE
		0.00	8.90	7.57%	102.60
Sales	Forecast	Error	ABS Error	% Error	Sq. Error
125	124	1	1	0.80%	1
153	152	1	1	0.65%	1
106	121	-15	15	14.15%	225
88	95	-7	7	7.95%	49
118	124	-6	6	5.08%	36
161	152	9	9	5.59%	81
133	121	12	12	9.02%	144
102	95	7	7	6.86%	49
138	124	14	14	10.14%	196
144	152	-8	8	5.56%	64
113	121	-8	8	7.08%	64
80	95	-15	15	18.75%	225
109	124	-15	15	13.76%	225
137	152	-15	15	10.95%	225
125	121	4	4	3.20%	16
109	95	14	14	12.84%	196
130	124	6	6	4.62%	36
165	152	13	13	7.88%	169
128	121	7	7	5.47%	49
96	95	1	1	1.04%	1

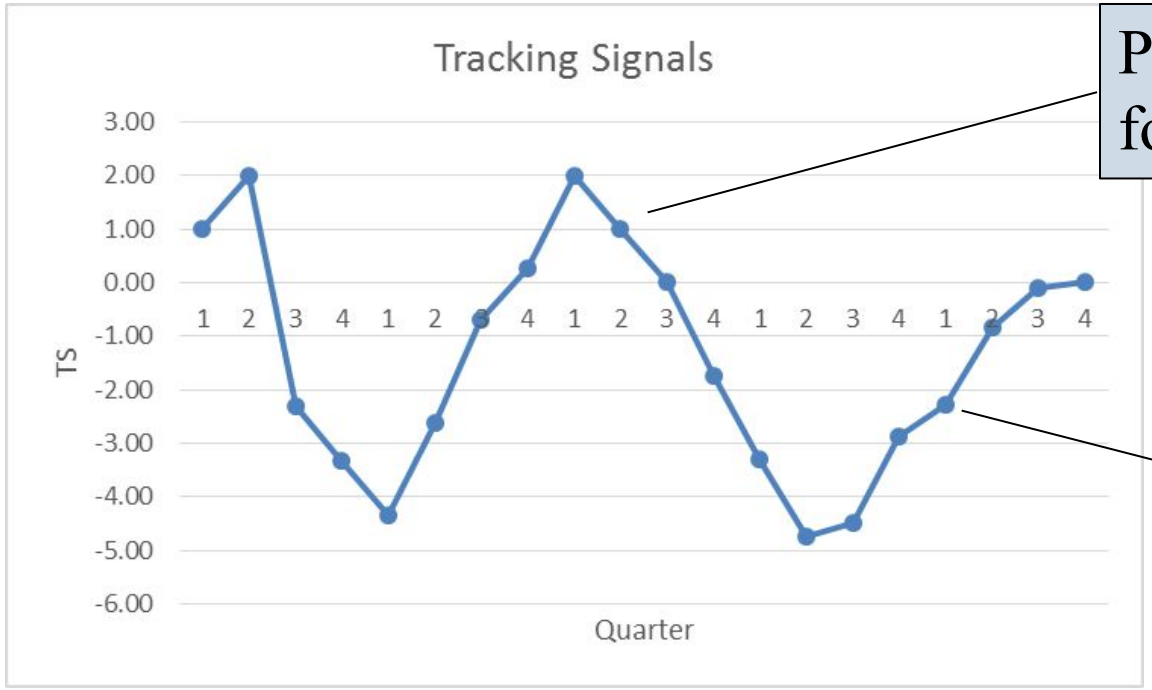
- Error = Sales – Forecast
- ABS Error = ABS(Error)
- % Error = ABS Error / Sales
- Sq. Error = Error^2
- ME = mean(Errors)
- MAE = mean(ABS Errors)
- MAPE = mean(% Errors)
- MSE = mean(Sq. Errors)

Tracking Signals (추적신호)

- Sum of Error = cumulated Error
- Sum of AE = cumulated ABS Error
- MAD = Sum of AE / Time Period
- TS = Sum of Error / MAD

Period	Error	ABS Error	Sum of Error	Sum of AE	MAE	TS
1	1	1	1.00	1.00	1.00	1.00
2	1	1	2.00	2.00	1.00	2.00
3	-15	15	-13.00	17.00	5.67	-2.29
4	-7	7	-20.00	24.00	6.00	-3.33
5	-6	6	-26.00	30.00	6.00	-4.33
6	9	9	-17.00	39.00	6.50	-2.62
7	12	12	-5.00	51.00	7.29	-0.69
8	7	7	2.00	58.00	7.25	0.28
9	14	14	16.00	72.00	8.00	2.00
10	-8	8	8.00	80.00	8.00	1.00
11	-8	8	0.00	88.00	8.00	0.00
12	-15	15	-15.00	103.00	8.58	-1.75
13	-15	15	-30.00	118.00	9.08	-3.31
14	-15	15	-45.00	133.00	9.50	-4.74
15	4	4	-41.00	137.00	9.13	-4.49
16	14	14	-27.00	151.00	9.44	-2.86
17	6	6	-21.00	157.00	9.24	-2.27
18	13	13	-8.00	170.00	9.44	-0.85
19	7	7	-1.00	177.00	9.32	-0.11
20	1	1	0.00	178.00	8.90	0.00

Tracking Signal Chart 추적신호차트



Positive TS - actual exceeds forecast 실제값이 예측값 초과

Negative TS - actual is less than forecast 실제값이 예측값보다 작음

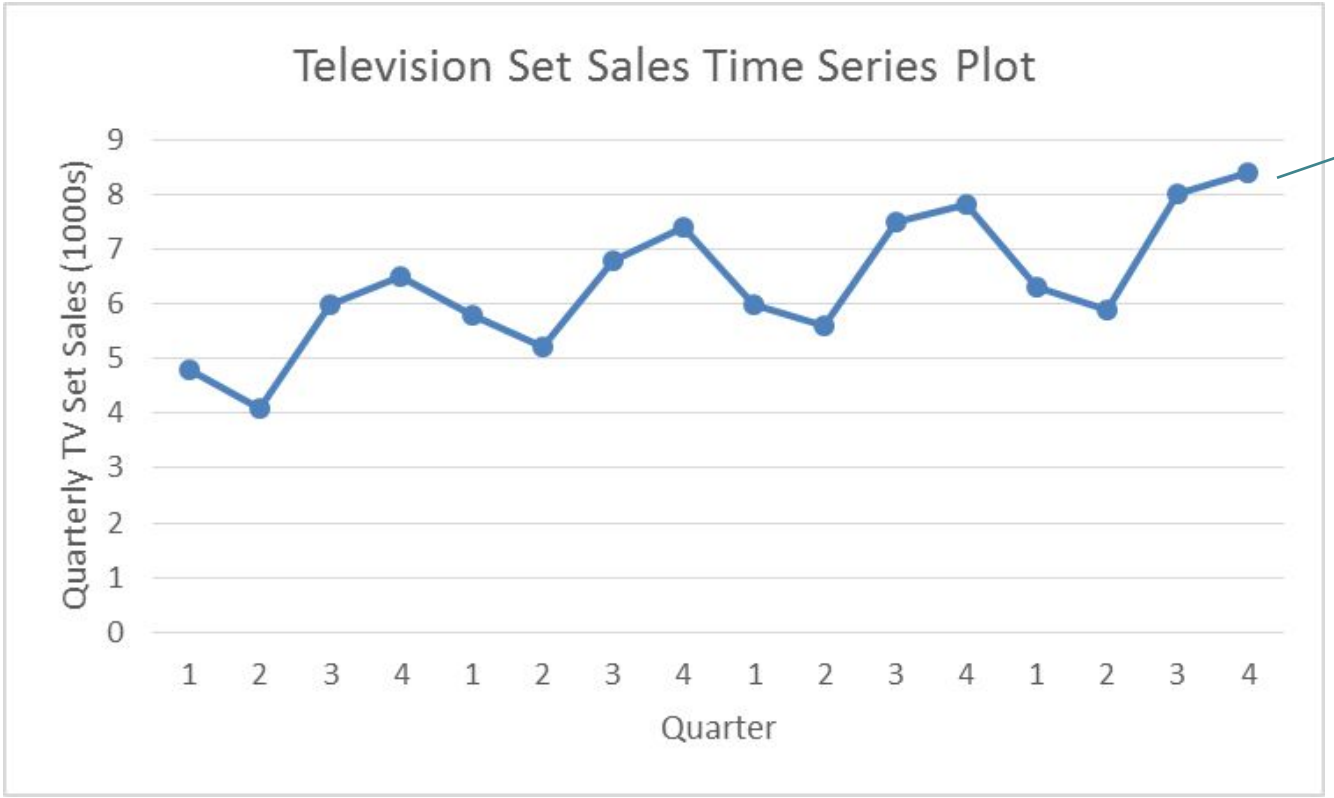
As long as TS is between -4 and 4, assume the model is working correctly. **-4와 4사이에 있으면 모델이 제대로 예측한다고 봄**

Television Set Sales Example
TV세일 계절성 추세예측 예제

Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8
	4	8.4

The table shows the television set sales for a particular manufacturer over the past four years. Forecast the quarterly television set sales for 5th year. **지난 4년간 특정 제조업자의 티비세트 세일을 보여주고 있습니다. 내년 (Year 5) 이 됐을때 쿼터별 세일을 예측해 보십시요.**

Time Series Plot (시계열차트)



The plot contains both a seasonal effect and a linear trend. 계절성과 추세성이 있음

Multiple Regression with Seasonality and Trend 다중회귀분석

Where

$\hat{Y}_t$ = estimate of forecast of sales in period t

$Qtr1$ = 1 if time period t corresponds to the first quarter of the year; 0 otherwise

$Qtr2$ = 1 if time period t corresponds to the second quarter of the year; 0 otherwise

$Qtr3$ = 1 if time period t corresponds to the third quarter of the year; 0 otherwise

t = time period

Time Series with Dummy Variables and Time Period 가변수

Year	Quarter	Sales (1000s)	Qrt1	Qtr2	Qtr3	Period
1	1	4.8	1	0	0	1
	2	4.1	0	1	0	2
	3	6	0	0	1	3
	4	6.5	0	0	0	4
2	1	5.8	1	0	0	5
	2	5.2	0	1	0	6
	3	6.8	0	0	1	7
	4	7.4	0	0	0	8
3	1	6	1	0	0	9
	2	5.6	0	1	0	10
	3	7.5	0	0	1	11
	4	7.8	0	0	0	12
4	1	6.3	1	0	0	13
	2	5.9	0	1	0	14
	3	8	0	0	1	15
	4	8.4	0	0	0	16

- Qtr1 = 1 if Quarter 1; 0 otherwise
- Qtr2 = 1 if Quarter 2; 0 otherwise
- Qtr3 = 1 if Quarter 3; 0 otherwise

Estimated Regression Equation
추정회귀식

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.98806594							
R Square	0.976274301							
Adjusted R Square	0.967646775							
Standard Error	0.216663753							
Observations	16							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	21.248	5.312	113.1580731	7.37582E-09			
Residual	11	0.516375	0.046943182					
Total	15	21.764375						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6.06875	0.162497815	37.34665609	6.12289E-13	5.711094721	6.426405279	5.711094721	6.426405279
Qrt1	-1.363125	0.157454336	-8.657271914	3.05975E-06	-1.709679657	-1.016570343	-1.709679657	-1.016570343
Qtr2	-2.03375	0.155107642	-13.111862	4.65532E-08	-2.375139619	-1.692360381	-2.375139619	-1.692360381
Qtr3	-0.304375	0.153682427	-1.980545245	0.073201043	-0.642627741	0.033877741	-0.642627741	0.033877741
Period	0.145625	0.012111872	12.0233272	1.14029E-07	0.118966949	0.172283051	0.118966949	0.172283051

$$Sales = 6.07 - 1.36\,Qtr1 - 2.03\,Qtr2 - .304\,Qtr3 + .146t$$

Forecast Sales for 5th Year 예측값

	A	B	C	D	E	F	G	H	I
1	Year	Quarter	Sales (1000s)	Qrt1	Qrt2	Qrt3	Period	Forecast	
18	5	1		1	0	0	17	7.1813	
19		2		0	1	0	18	6.6563	
20		3		0	0	1	19	8.5313	
21		4		0	0	0	20	8.9813	
22									
23	SUMMARY OUTPUT								
24									
25	*Regression Statistics*								
26	Multiple R	0.9880659							
27	R Square	0.9762743							
28	Adjusted R Square	0.9676468							
29	Standard Error	0.2166638							
30	Observations	16							
31									
32	ANOVA								
33		*df*	*SS*	*MS*					
34	Regression	4	21.248	5.312	113				
35	Residual	11	0.516375	0.0469432					
36	Total	15	21.764375						
37									
38		*Coefficients*	*Standard Error*	*t Stat*	*P-*				
39	Intercept	6.06875	0.162497815	37.346656	6.1				
40	Qrt1	-1.363125	0.157454336	-8.657272	3.(				
41	Qrt2	-2.03375	0.155107642	-13.11186	4.6				
42	Qrt3	-0.304375	0.153682427	-1.980545	0.				
43	Period	0.145625	0.012111872	12.023327	1.				

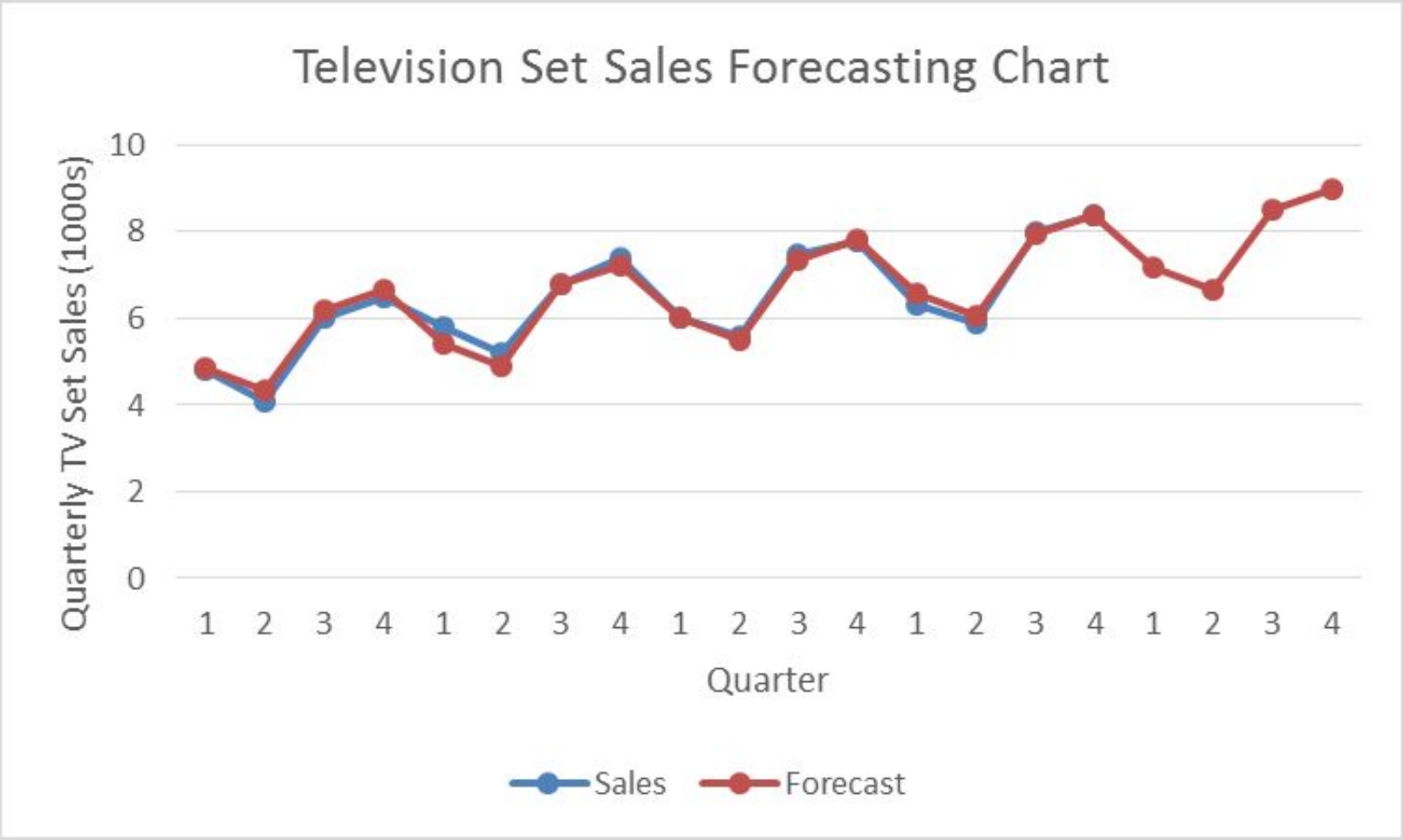
$$= 6.07 - 1.36 * 1 - 2.03 * 0 - .304 * 0 + .146 * 17$$

$$= 6.07 - 1.36 * 0 - 2.03 * 1 - .304 * 0 + .146 * 18$$

$$= 6.07 - 1.36 * 0 - 2.03 * 0 - .304 * 1 + .146 * 19$$

$$= 6.07 - 1.36 * 0 - 2.03 * 0 - .304 * 0 + .146 * 20$$

Forecast Chart 예측차트



Forecast Errors 예측에러

		ME	MAE	MAPE	MSE
		0.00	0.14	0.02	0.03
Sales (1000s)	Forecast	Error	ABS Error	% Error	Sq. Error
4.8	4.8513	-0.05	0.05	0.01	0.00
4.1	4.3263	-0.23	0.23	0.06	0.05
6	6.2013	-0.20	0.20	0.03	0.04
6.5	6.6513	-0.15	0.15	0.02	0.02
5.8	5.4338	0.37	0.37	0.06	0.13
5.2	4.9088	0.29	0.29	0.06	0.08
6.8	6.7838	0.02	0.02	0.00	0.00
7.4	7.2338	0.17	0.17	0.02	0.03
6	6.0163	-0.02	0.02	0.00	0.00
5.6	5.4913	0.11	0.11	0.02	0.01
7.5	7.3663	0.13	0.13	0.02	0.02
7.8	7.8163	-0.02	0.02	0.00	0.00
6.3	6.5988	-0.30	0.30	0.05	0.09
5.9	6.0738	-0.17	0.17	0.03	0.03
8	7.9488	0.05	0.05	0.01	0.00
8.4	8.3988	0.00	0.00	0.00	0.00

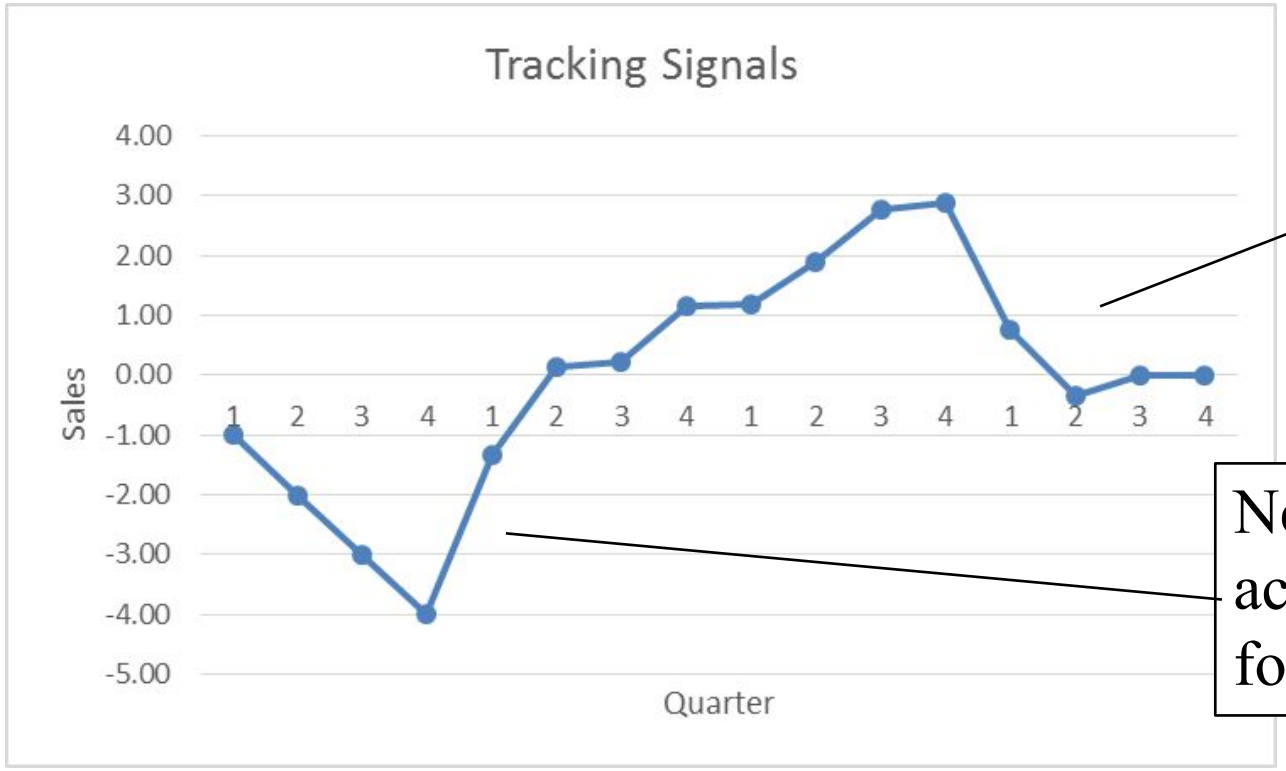
- Error = Sales – Forecast
- ABS Error = ABS(Error)
- % Error = ABS Error / Sales
- Sq. Error = Error^2
- ME = mean(Errors)
- MAE = mean(ABS Errors)
- MAPE = mean(% Errors)
- MSE = mean(Sq. Errors)

Tracking Signals 추적신호

Period	Error	ABS Error	Sum of Error	Sum of AE	MAE	TS
1	-0.05	0.05	-0.05	0.05	0.05	-1.00
2	-0.23	0.23	-0.28	0.28	0.14	-2.00
3	-0.20	0.20	-0.48	0.48	0.16	-3.00
4	-0.15	0.15	-0.63	0.63	0.16	-4.00
5	0.37	0.37	-0.26	1.00	0.20	-1.32
6	0.29	0.29	0.03	1.29	0.21	0.13
7	0.02	0.02	0.04	1.30	0.19	0.23
8	0.17	0.17	0.21	1.47	0.18	1.14
9	-0.02	0.02	0.19	1.49	0.17	1.17
10	0.11	0.11	0.30	1.60	0.16	1.90
11	0.13	0.13	0.44	1.73	0.16	2.78
12	-0.02	0.02	0.42	1.75	0.15	2.89
13	-0.30	0.30	0.12	2.04	0.16	0.77
14	-0.17	0.17	-0.05	2.22	0.16	-0.33
15	0.05	0.05	0.00	2.27	0.15	-0.01
16	0.00	0.00	0.00	2.27	0.14	0.00

- Sum of Error = cumulated Error
- Sum of AE = cumulated ABS Error
- MAD = Sum of AE / Time Period
- TS = Sum of Error / MAD

Tracking Signal Chart 추적신호차트



Positive TS - actual exceeds forecast

Negative TS - actual is less than forecast

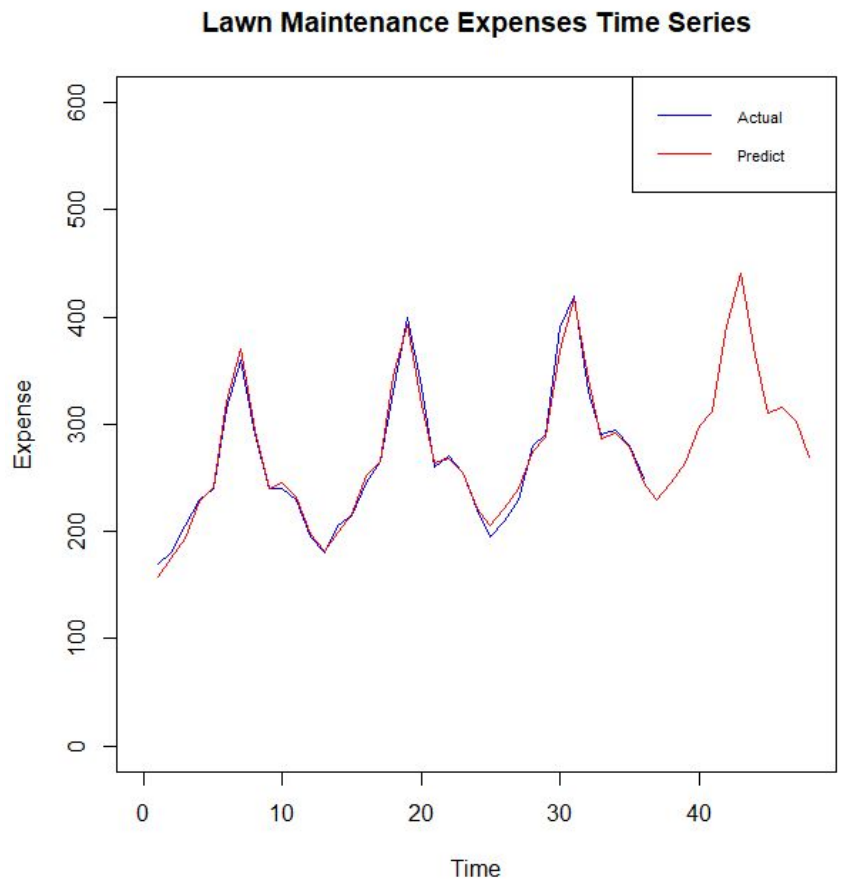
As long as TS is between -4 and 4, assume the model is working correctly.

Lawn-Maintenance Expense
잔디유지비용 계절추세 예측예제

Month	Year 1	Year 2	Year 3
January	170	180	195
February	180	205	210
March	205	215	230
April	230	245	280
May	240	265	290
June	315	330	390
July	360	400	420
August	290	335	330
September	240	260	290
October	240	270	295
November	230	255	280
December	195	220	250

- Three years of monthly law-maintenance expenses ($) for a six-unit apartment house in southern Florida. **남플로리다에 있는 6개 아파트의 지난 3년간의 유지비용을 보여주고 있습니다. 시계열차트를 그리고, 내년 (Year 4)의 월별 유지비용을 예측하십시요.**

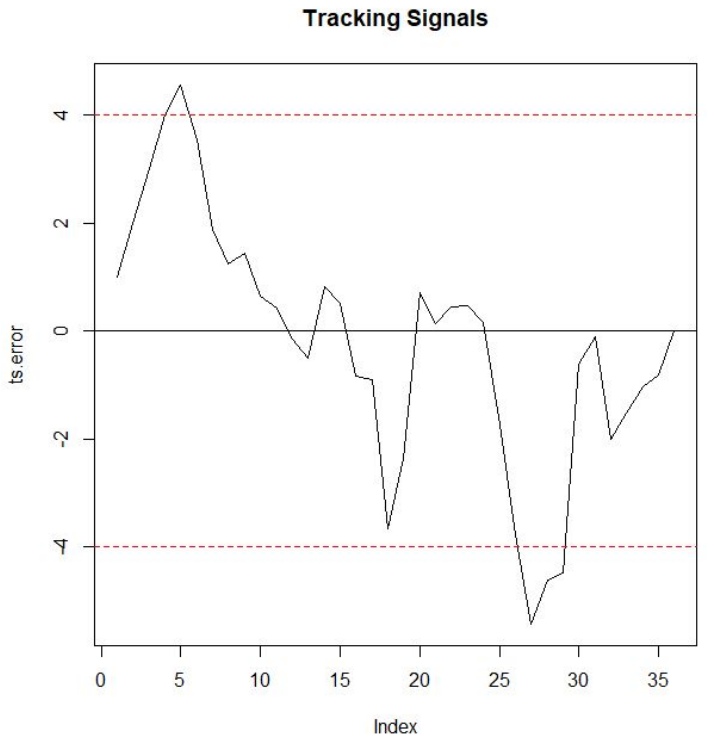
Forecasting Sales for 4th Year 예측값




```
month1  228.757
month2  245.424
month3  263.757
month4  298.757
month5  312.091
month6  392.091
month7  440.425
month8  365.425
month9  310.425
month10 315.426
month11 302.092
month12 268.759
```

Forecast Errors

ME = 0
MAE = 5.89
MAPE = 2.26
MSE = 61.08



Exercise #9
Umbrella Sales Example

- Plot the time series 시계열차트
- Conduct a seasonal forecast and draw the forecasting chart
  계절예측수행하고 예측차트
- Forecast year 6 6년 예측
- Create an error table and calculate ME, MAE, MAPE, and MSE
  에러테이블과 에러측정치
- Create a tracking signal table and plot the signals 추적신호와 차트

Exercise #9
TV Set Sales Example

- Plot the time series 시계열차트
- Conduct a seasonal forecast and draw the forecasting chart 계절예측과 그래프
- Forecast year 5 5년 예측
- Create an error table and calculate ME, MAE, MAPE, and MSE 에러테이블, 에러측정치
- Create a tracking signal table and plot the signals 추적신호, 그래프

Exercise #9
Lawn-Maintenance Expense Example

- Plot the time series 시계열차트
- Conduct a seasonal forecast and draw the forecasting chart 계절예측과 그래프
- Forecast year 4 4년 예측
- Create an error table and calculate ME, MAE, MAPE, and MSE 에러테이블, 에러측정치
- Create a tracking signal table and plot the signals 추적신호와 차트

Decomposition

분해법

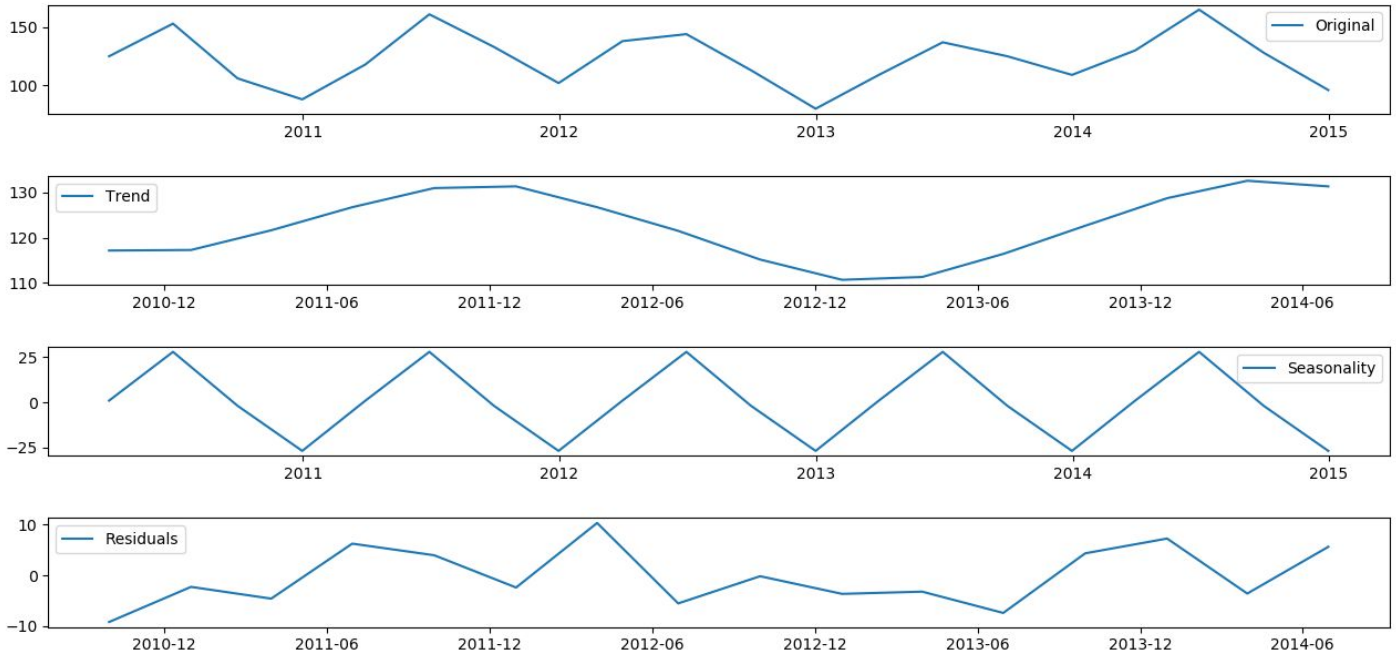
Decomposing Seasonal Data
계절요소의 분해

- Separating the time series into seasonal, trend, and random components. 시계열데이터를 계절, 추세, 무작위요소로 분리

```
from statsmodels.tsa.seasonal import
seasonal_decompose
decomposition = seasonal_decompose(ts)
trend = decomposition.trend
seasonal = decomposition.seasonal
residual = decomposition.resid
```

Decomposition Plot

```
plt.subplot(411)
plt.plot(ts, label='Original')
plt.legend(loc='best')
plt.subplot(412)
plt.plot(trend, label='Trend')
plt.legend(loc='best')
plt.subplot(413)
plt.plot(seasonal,label='Seasonality')
plt.legend(loc='best')
plt.subplot(414)
plt.plot(residual, label='Residuals')
plt.legend(loc='best')
plt.tight_layout()
```



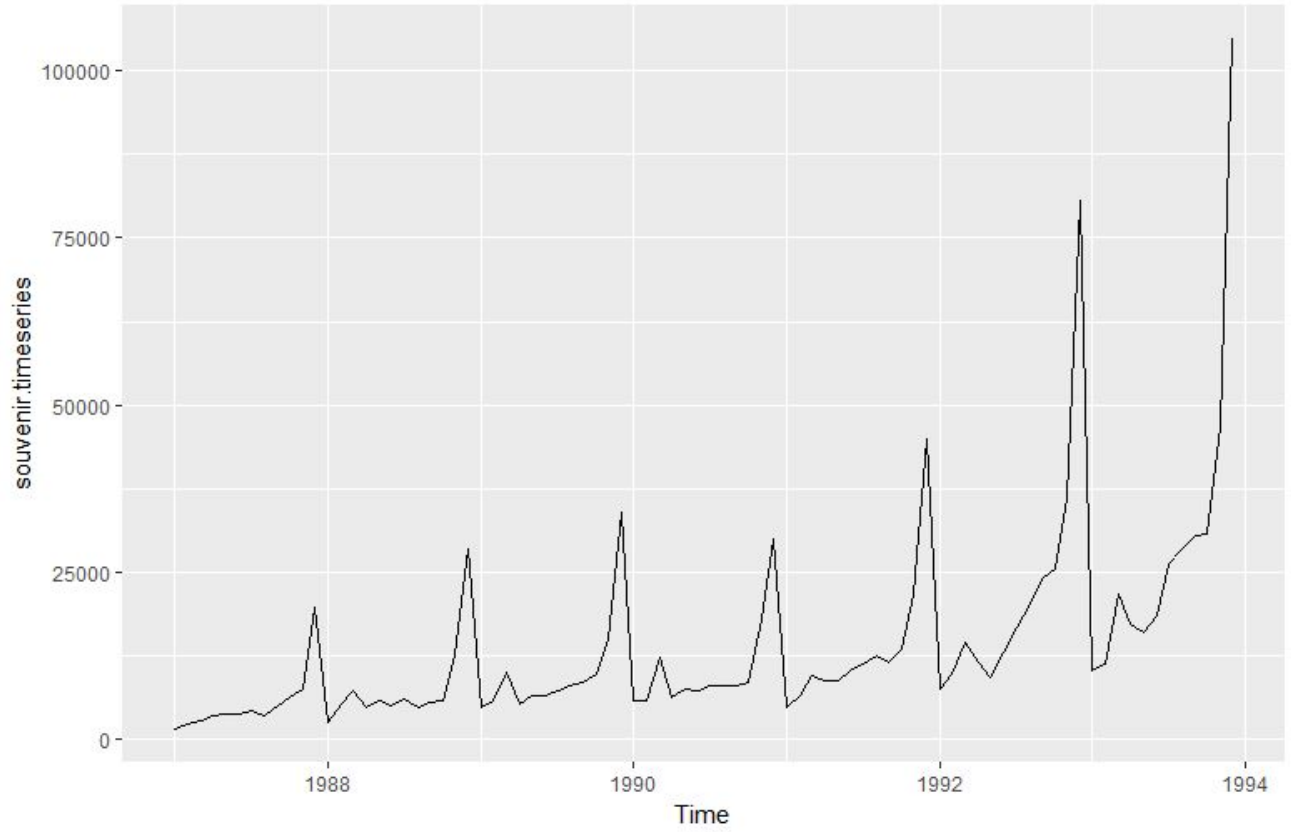
Souvenir Shop Sales Example
기념품가게 승법계절예측

The file below contains monthly sales for a souvenir shop at a beach resort town in Queensland, Australia, for January 1987-December 1993. Forecast the monthly sales for 1994. **아래 파일은 오스트렐리아 퀸즈랜드해변 리조트타운에 있는 기념품 가게의 월별 세일내역으로 1987년 12월부터 1993년 12월까지의 매출량이 기록되어 있습니다. 1994년의 월별 매출을 예측하십시요.**

http://robjhyndman.com/tsdldata/data/fancy.dat

Multiplicative Model 승법모델

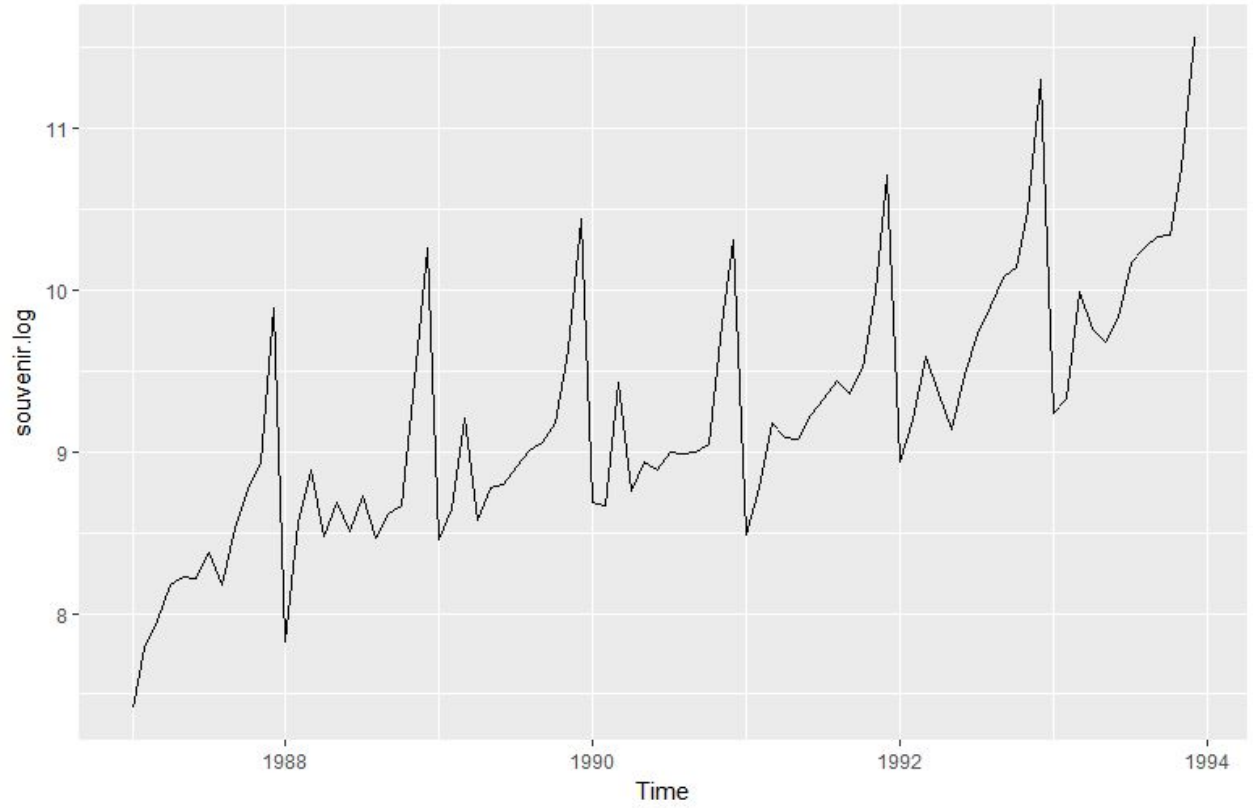
- The size of the seasonal fluctuations and random fluctuations seem to increase with the level of the time series. 시간이 지남에 따라 추세, 계절성이 증폭됨



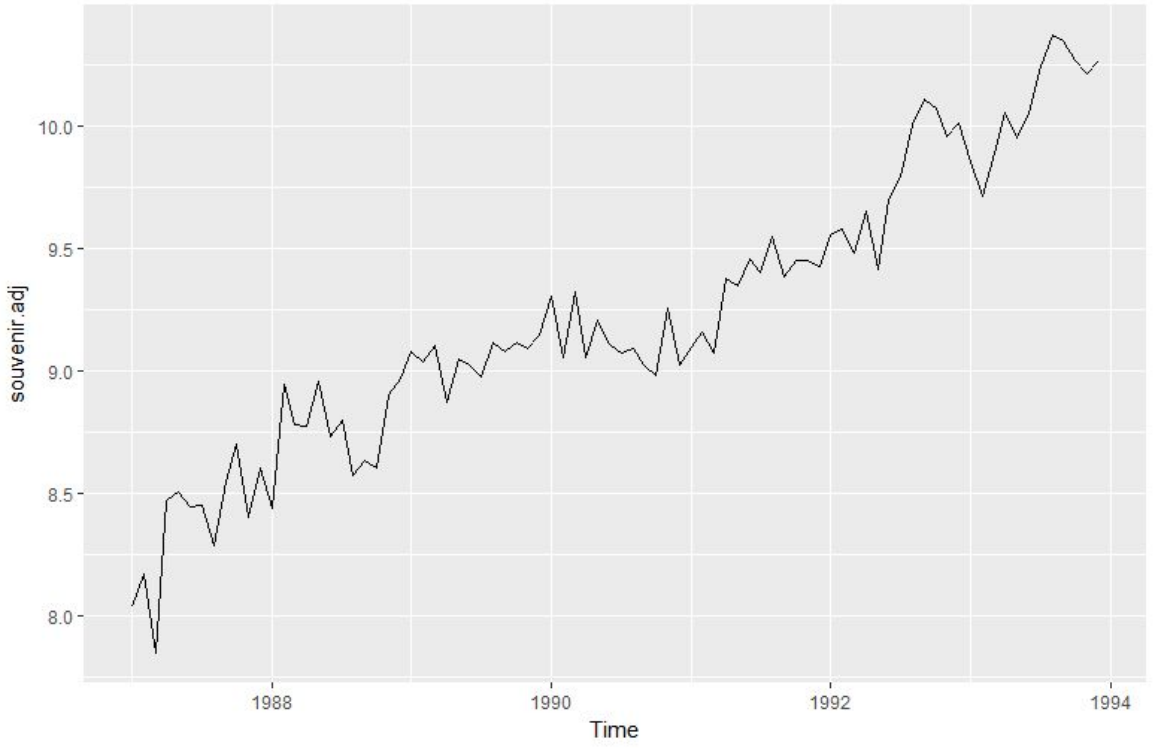
Log Transformation (로그변환)

- The log-transformed time series can be described using an additive model. **로그변환을 하고나면 가법모델로 바뀜**

```
souvenior_log =
np.log(souvenior)
plt.plot(souvenior_log)
```

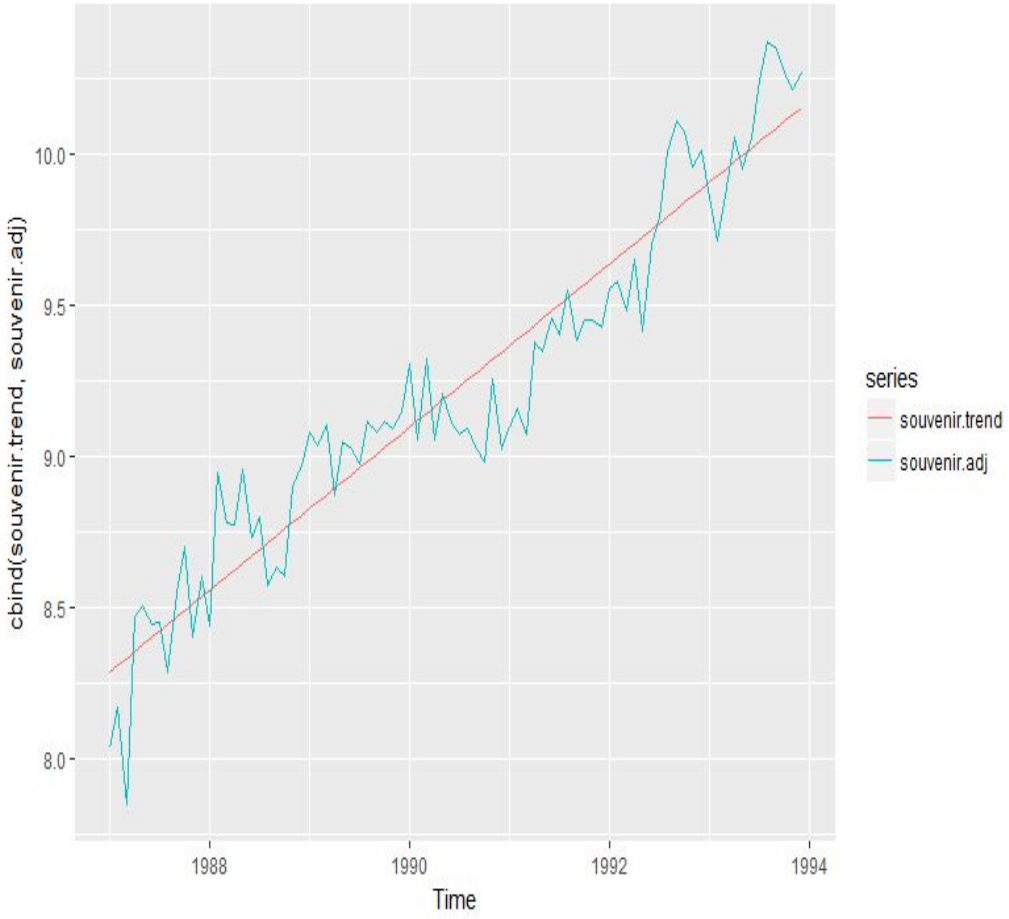


Seasonal Decomposition 계절분해



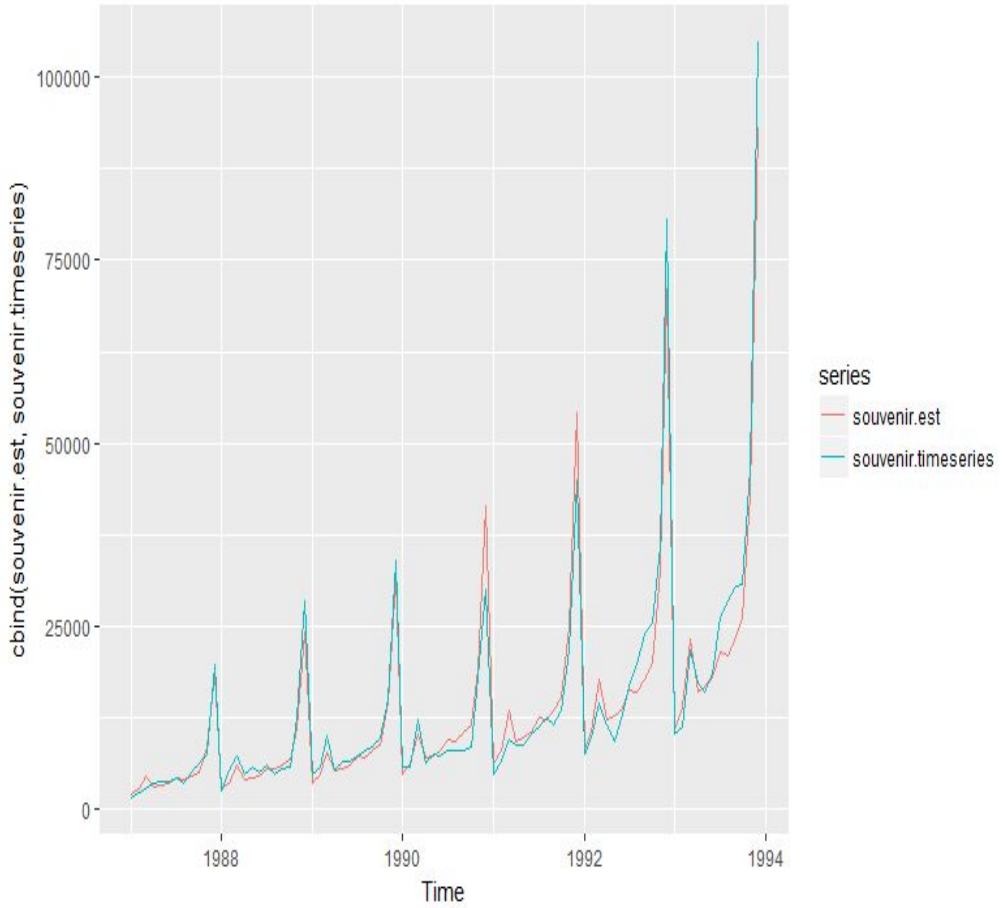

```
from
statsmodels.tsa.seasonal
import seasonal_decompose

decomp =
seasonal_decompose(souven
ior_log)

Seasonal =
decomp.seasonal

souvenior_adj =
souvenior_log - seasonal
```

Linear Regression 추세선




```
import statsmodels.api as sm
t = np.arange(1,85)
t = sm.add_constant(t)
model = sm.OLS(y,x).fit()
model.summary()
souvenior_trend = model.predict(t)
souvenior_trend =
pd.DataFrame(souvenior_trend,
index=souvenior.index,
columns=['sales'])
plt.plot(souvenior_adj)
plt.plot(souvenior_trend)
```

Re-composition 복구




```
souvenior_comp =
souvenior_trend + seasonal

souvenior_pred =
np.exp(souvenior_comp)

plt.plot(souvenior)

plt.plot(souvenior_pred)
```

Exercise #9
Souvenir Example

- Import the souvenir data and transform using log.
  기념품 데이터를 가져온후 로그변환
- Decompose the seasonality and conduct the regression analysis. 계절성분해후 회귀분석
- Recompose the data and forecast sales for next 12 months. 데이터를 복귀하여 다음 12개월예측

Exercise #9
Souvenir Example

- Create an error table and calculate ME, MAE, MAPE, and MSE 에러테이블, 측정치
- Create a tracking signal table and plot the signals 추적신호와 차트
- Compare this result to the result from other forecast methods. 다른 예측법의 결과와 비교

Exercise #9
Umbrella Sales Example

- Create an umbrella time series (freq='Q') and plot it using plt.plot **분기별 데이터로 만든후 그래프**
- Forecast year 6 using seasonal decomposition method (subtract seasonal factor linear regression add seasonal factor)
  시계열데이터를 분해하여 6년을 예측
- Compare this result to the result from multiple regression
  다중회귀의 결과와 비교

Exercise #8
TV Sets Sales Example

- Create a tvset time series (freq='Q') and plot it using plt.plot
  분기별 데이터로 만든후 그래프
- Forecast year 5 using seasonal decomposition method (subtract seasonal factor linear regression add seasonal factor)
  시계열데이터를 분해하여 5년을 예측
- Compare this result to the result from from multiple regression
  다중회귀의 결과와 비교

Exercise #8
Lawn-Maintenance Expense Example

- Create a tvset time series (freq='M') and plot it using plt.plot
  월별 데이터로 만든후 그래프
- Forecast year 4 using seasonal decomposition method (subtract seasonal factor linear regression add seasonal factor)
  시계열데이터를 분해하여 4년을 예측
- Compare this result to the result from from multiple regression
  다중회귀의 결과와 비교

Exercise #9

- Clothing Store Sales Example
  - http://www.census.gov/retail/ □ Monthly Retail Trade Report □Time Series/Trend Charts: Create your own customizable time series □ Clothing store (4481) □ Clothing store sales from 2010 to 2018. Forecast monthly sales in 2019. **미국인구조사국에 가서 2010년부터 2018년 옷가게 세일데이터를 받은후 2019년 월별세일 예측**
- Furniture Store Sales Example
  - Find furniture store sales from 2010 to 2018 from the US Census and forecast monthly sales in 2019 **2010년부터 2018년까지 가구점세일 데이터를 받은후 2019년 월별 가구세일 예측**

Clothing Store Sales from US Census
옷가게 세일 예측

TIME SERIES / TREND CHARTS

Please follow the numbers in order.

1 Select the report/survey from which you wish to retrieve data:
Monthly Retail Trade and Food Services

2 Select a date range:
Start: 2010 End: 2012

3 Select Industry or Category:
4481: Clothing Stores

4 Select one Item :
Sales - Monthly

5 Select Geographical Level:
U.S. Total

Select as available:
- ☐ Seasonally Adjusted
- ☑ Not Seasonally Adjusted
- ☐ Show Estimates of Sampling Variability

GET DATA

Monthly clothing store sales from 2010 to 2012.

Source: Mon etail Trade and Food Services (Definitions)
4481: Clothi tores: U.S. Total — Not Seasonally Adjusted Sa
TXT XLS-V XLS-H Bar Chart Line Chart

Year	Jan	Feb	Mar	Apr
2010	9,931	10,605	13,174	12,951
2011	10,201	11,407	13,760	13,912
2012	10,752	12,720	15,342	14,148

Time Series Models
시계열모델의 종류

Smoothing Models

- Simple Moving Average
- Weighted Moving Average
- Exponential Smoothing

Trend Models

- Linear
- Quadratic
- Exponential
- Auto regression

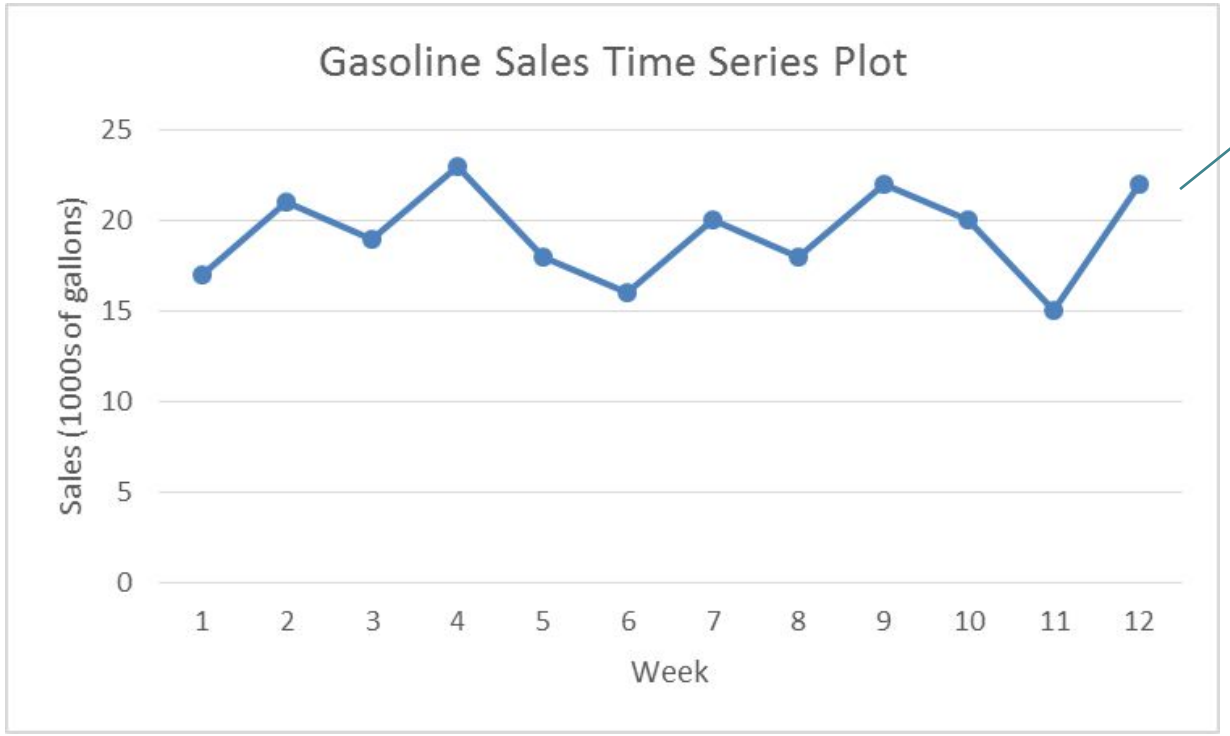
Gasoline Sales Example
가솔린세일 비선형추세 예측

Week	Sales (1000s of gallons)
1	17
2	21
3	19
4	23
5	18
6	16
7	20
8	18
9	22
10	20
11	15
12	22

The table shows the number of gallons of gasoline sold by a gasoline distributor in Bennington, Vermont, over the past 12 weeks. Forecast the gasoline sales for week 13.

다음 테이블은 지난 12주동안 버몬트 베닝톤에 있는 가솔린 유통업자에 의해서 팔린 가솔린의 양 (갤론단위)을 보여주고 있습니다. 13주가 됐을때 가솔린 세일을 예측하십시요.

Time Series Plot 시계열차트



A horizontal pattern is present!

The data fluctuate around the sample mean of 19,250 gallons. 특정값 주의에서 변동됨

Three-Week Moving Average
3주 이동평균

Week	Sales (1000s of gallons)	3PMA
1	17	
2	21	
3	19	
4	23	19
5	18	21
6	16	20
7	20	19
8	18	18
9	22	18
10	20	20
11	15	20
12	22	19
13		19

- The average of the most recent three data values in the time series as the forecast for the next period
  최근 3주를 평균냄
- Forecast for Week 13
  = (20 +15 + 22)/3 = 19

Six-Week Moving Average
6주 이동평균

Week	Sales (1000s of gallons)	6PMA
1	17	
2	21	
3	19	
4	23	
5	18	
6	16	
7	20	19.00
8	18	19.50
9	22	19.00
10	20	19.50
11	15	19.00
12	22	18.50
13		19.50

- An equal weight is placed on each value that is being averaged. 가중치가 같음
- Forecast for Week 13 = (20 +18+22+20+15+22)/6 = 19.5

Moving Average Forecast of Order k
k차의 이동평균예측

- $$\mathbf{F_{t+1}} = \frac{\sum(\textbf{most recent k data values})}{\mathbf{k}}$$

$$= \frac{\mathbf{Y_t + Y_{t-1} + \ \ldots + Y_{t-k+1}}}{\mathbf{k}}$$

Where

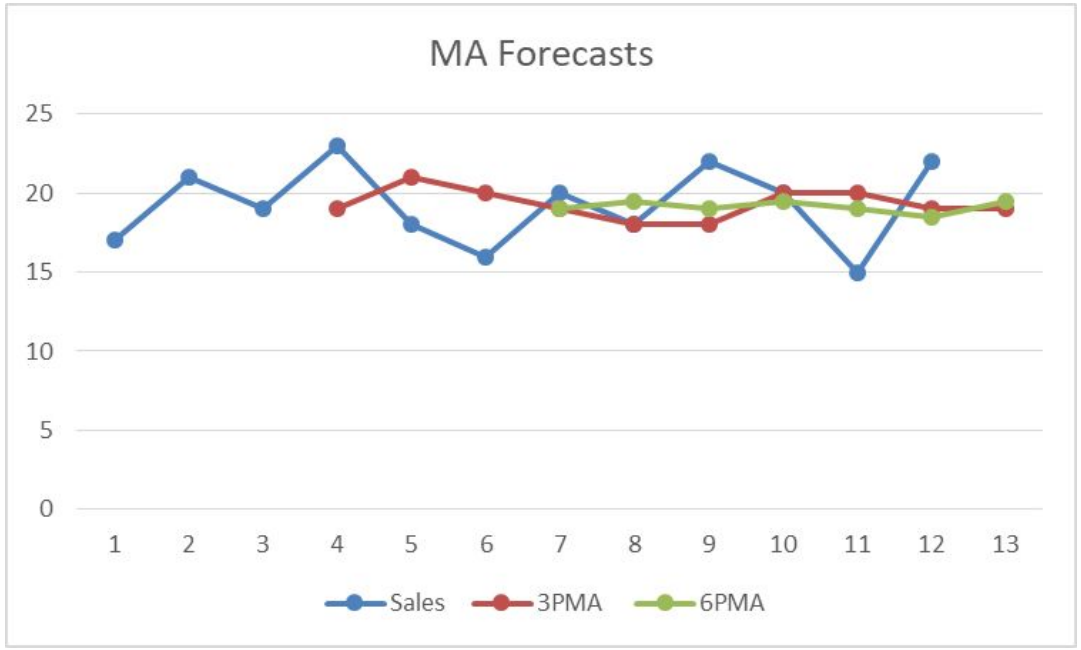
F_{t+1} = forecast of the times series for period t + 1

Y_t = actual value of the time series in period t

Order Selection 차수선택

- You must first select the order, or number of time series values, to be included in the moving average
  이동평균을 낼때 우선 평균낼 기간을 선택
- Use trial and error to determine the value of k that minimizes MSE. 평균오차제곱값을 최소화 할수있는 기간을 여러번의 시도로 결정

Moving Average Chart 이동평균표



Longer gives more smoothing. 기간이 길수록 더 스무드한 예측

Shorter reacts quicker to trends. 시간이 짧을수록 더 트렌드를 잘 반영

The most accurate moving average forecasts of gasoline sales can be obtained using a moving average of order k=6 with MSE =6.79
기간이 6일때 오차가 가장 작음.

.rolling Function

- Can be applied on a series of data.
- Specify the **window=n** argument and apply the appropriate statistical function on top of it

```
gas = np.array([17, 21, 19, 23, 18, 16, 20, 18, 22,
20, 15, 22])
t = range(1, 13)
df = pd.DataFrame({'t':t, 'gas':gas})
df.set_index('t', inplace=True)
df.rolling(window=3).mean()
```

Three-Week Weighted Moving Average 3주 가중이동평균

Week	Sales (1000s of gallons)	Wt.	WMA
1	17	0.166667	
2	21	0.333333	
3	19	0.5	
4	23		19.33
5	18		21.33
6	16		19.83
7	20		17.83
8	18		18.33
9	22		18.33
10	20		20.33
11	15		20.33
12	22		17.83
13			19.33

- The weighted average of the most recent three values as the forecast **최근 3주동안 가중평균**
- Forecast for Week 13 = (3/6)*22 + (2/6)*15 + (1/6)*20 = 19.33

Weighted Moving Average
가중이동평균

$$F_{t+1} = \sum w_t A_t$$

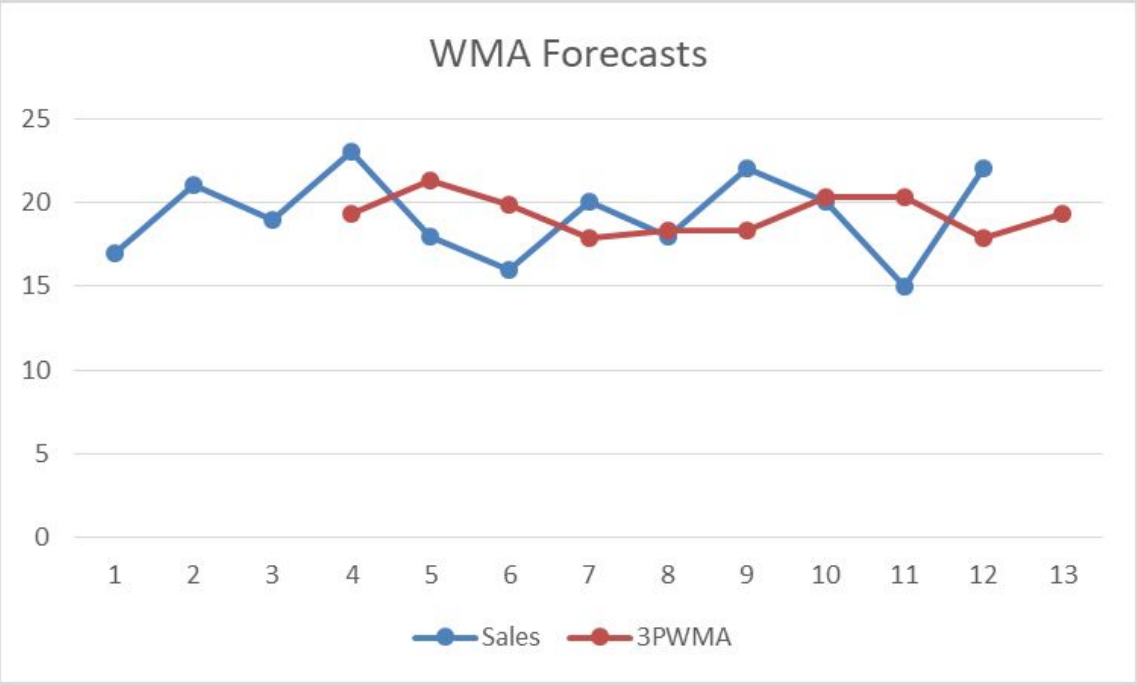
Where

F_{t+1} = forecast of the times series for period t + 1

A_t = actual value of the time series in period t

- All the weights must sum to one.
- The weighted moving average permits an unequal weighting on prior time periods.

Weighted Moving Average Chart
가중이동평균차트



- More responsive to trends because of usually more weight on recent data!
  일반적으로 최근 데이터에 더 가중치를 두기때문에 더 변화를 잘 반영

Weight Selection (가중치 선택)

- Use trial and error to determine the number of data values and weights. **이것도 여러번의 시도로 가중치를 결정**
- If the recent past is a better predictor of the future than the distant past, larger weights should be given to the more recent observations. **가까운 년도가 더 영향력이 있는 경우가 많아서 최근것에 더 가중치를 둠**
- When the time series is highly variable, selecting approximately equal weights for the data values may be best. **너무 변동이 많은 경우는 같은 가중치를 줌**
- Use the combination of number of data values and weights that minimizes MSE! **MSE를 최소로 하는 값을 찾음**

Exponential Smoothing 지수평활

Week	Sales (1000s of gallons)	alpha	Exp
1	17	0.2	
2	21		17.00
3	19		17.80
4	23		18.04
5	18		19.03
6	16		18.83
7	20		18.26
8	18		18.61
9	22		18.49
10	20		19.19
11	15		19.35
12	22		18.48
13			19.18

- The weighted average of actual value in period 12 and the forecast for period 12. **지난기간의 실제값과 예측값의 가중평균**
- Forecast for Week 13 =.2*22 + (1-.2)*18.48
  = 18.48 + (22–18.48)*.2 = 19.18
- The forecast for week2 equals the actual value of the time series in week1 (naïve method). **처음 예측값은 일주의 실제값으로 함**

Exponential Smoothing Forecast
지수평활예측

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

  Where

  F_{t+1} = forecast of the time series for period t + 1

  Y_t = actual value of the time series in period t

  F_t = forecast of the time series for period t

  α = smoothing constant ($0 \leq \alpha \leq 1$)

- Need just three pieces of data to start: last period's forecast, last period's actual value, smoothing coefficient, α.

Alpha Selection 알파의 선택

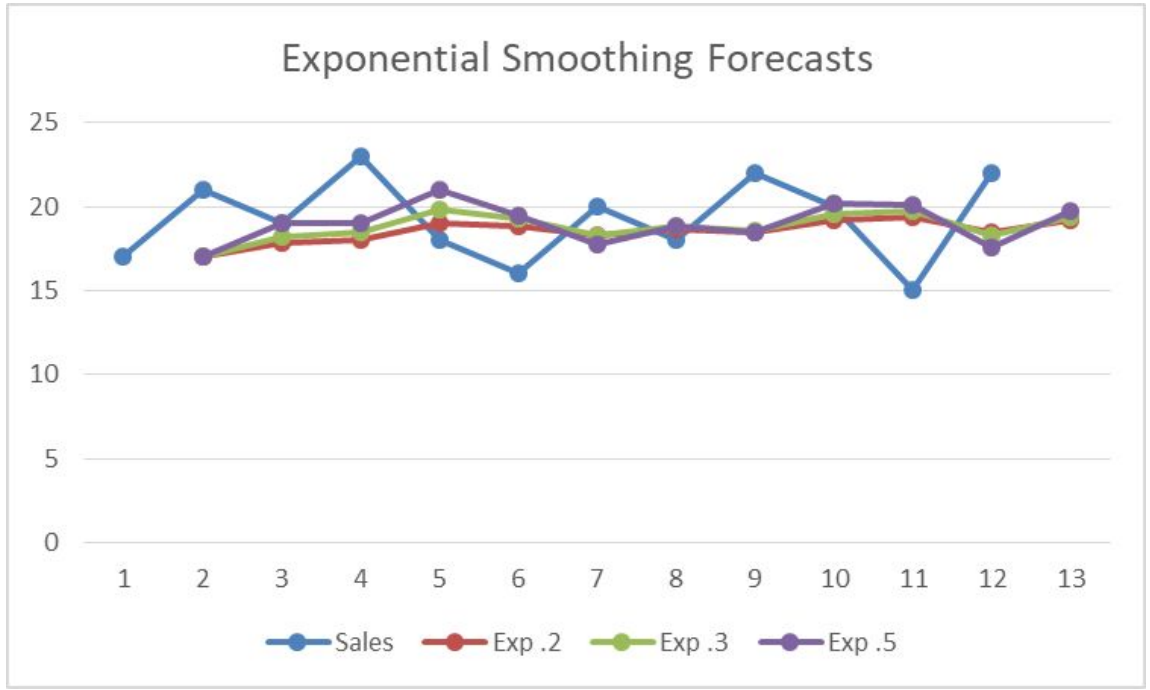
- Use trial and error to determine the value of alpha minimizes the MSE! **여러번의 시도로 MSE를 최소화하는 알파값을 구함**
- Larger values of the smoothing constant allows the forecast to react more quickly to changing conditions. **숫자가 클수록 변화에 더 잘 반응하는 예측을 할수있음**

α = 0.05-0.1, relatively stable
α = 0.15-0.3, rapid growth

Different Alpha Values
알파값의 비교

		Forecast					
Week	Sales (1000s of gallons)	alpha	Exp	alpha	Exp	alpha	Exp
1	17	0.2		0.3		0.5	
2	21		17.00		17.00		17.00
3	19		17.80		18.20		19.00
4	23		18.04		18.44		19.00
5	18		19.03		19.81		21.00
6	16		18.83		19.27		19.50
7	20		18.26		18.29		17.75
8	18		18.61		18.80		18.88
9	22		18.49		18.56		18.44
10	20		19.19		19.59		20.22
11	15		19.35		19.71		20.11
12	22		18.48		18.30		17.55
13			19.18		19.41		19.78

Exponential Smoothing Chart
지수평활표



- Most frequently used method 가장 많이 사용되는 방법

.ewm Function

- Assigns the weights exponentially.
- Specify any of the com, span, **halflife** argument and apply the appropriate statistical function on top of it.

```
df.ewm(alpha=.2).mean()
df.ewm(alpha=.3).mean()
```

Exercise #9
Gasoline Sales Example

- Forecast week 13 using simple calculation, formulas, and forecast functions **단순계산, 공식, 예측함수등을 이용하여 13주를 예측**
  - Simple moving average (3 weeks, 6 weeks) **이동평균**
  - Weighted moving average (3/6, 2/6, 1/6) **가중이동평균**
  - Exponential smoothing method (alpha = .2, .3, .5) **지수평활**

Exercise #9

- Use Umbrella Sales, TV Sets Sales, and Lawn-Maintenance Expense to forecast next time period using simple calculation, formulas, and forecast functions **단순계산, 공식, 예측함수를 이용하여 예측**
  - Simple moving average (3 quarters or months, 6 quarters or months) **이동평균**
  - Weighted moving average (3/6, 2/6, 1/6) **가중이동평균**
  - Exponential smoothing method (alpha = .2, .3, .5) **지수평활**