Day 8

RNN/Forecasting

Data Preparation for RNN

- Load data
- Drop date
- Scaling (MinMaxScaler)
- Split x and y (with appropriate time steps)
- Split train and test (80%, 2/3, etc.)
- Reshape an input to 3D with the format of [samples, time steps and features]

LSTM Inputs

- Must be 3D 3차원이어야 함
- The 3 input dimensions are: samples, time steps, and features. 3차원은 샘플의 수, 타임스텝의 수, 변수의수임
- The input_shape argument on the first hidden layer takes the number of time steps and features. 첫번째 히든레이어의 입력모양은 타임스텝의 수와 변수의 수를 받음

Exercise #8

- Collect the Google stock price from yahoo finance and take the Close price for the analysis
  - January 2012 to 2019
- Check the null values df.isna().sum()
- Scaling (MinMaxScaler between 0 and 1)
- Preparing x and y with 60 timesteps
- Train and test split
- Reshaping(x_train.shape[0], x_train.shape[1], 1)

Index	Open	High	Low	Close	Adj Close	Volume
2012-01-03...	326.797	334.409	326.512	333.038	333.038	7345600
2012-01-04...	332.848	335.46	330.641	334.474	334.474	5722200
2012-01-05...	331.396	332.317	328.443	329.835	329.835	6559200
2012-01-06...	329.905	330.33	325.22	325.335	325.335	5380400
2012-01-09...	323.574	323.824	310.926	311.542	311.542	11633500
2012-01-10...	315.19	317.217	308.764	311.882	311.882	8782400
2012-01-11...	312.062	315.01	310.871	313.293	313.293	4795200
2012-01-12...	315.926	316.762	313.564	315.135	315.135	3746600
2012-01-13...	313.443	313.789	310.841	312.808	312.808	4609900

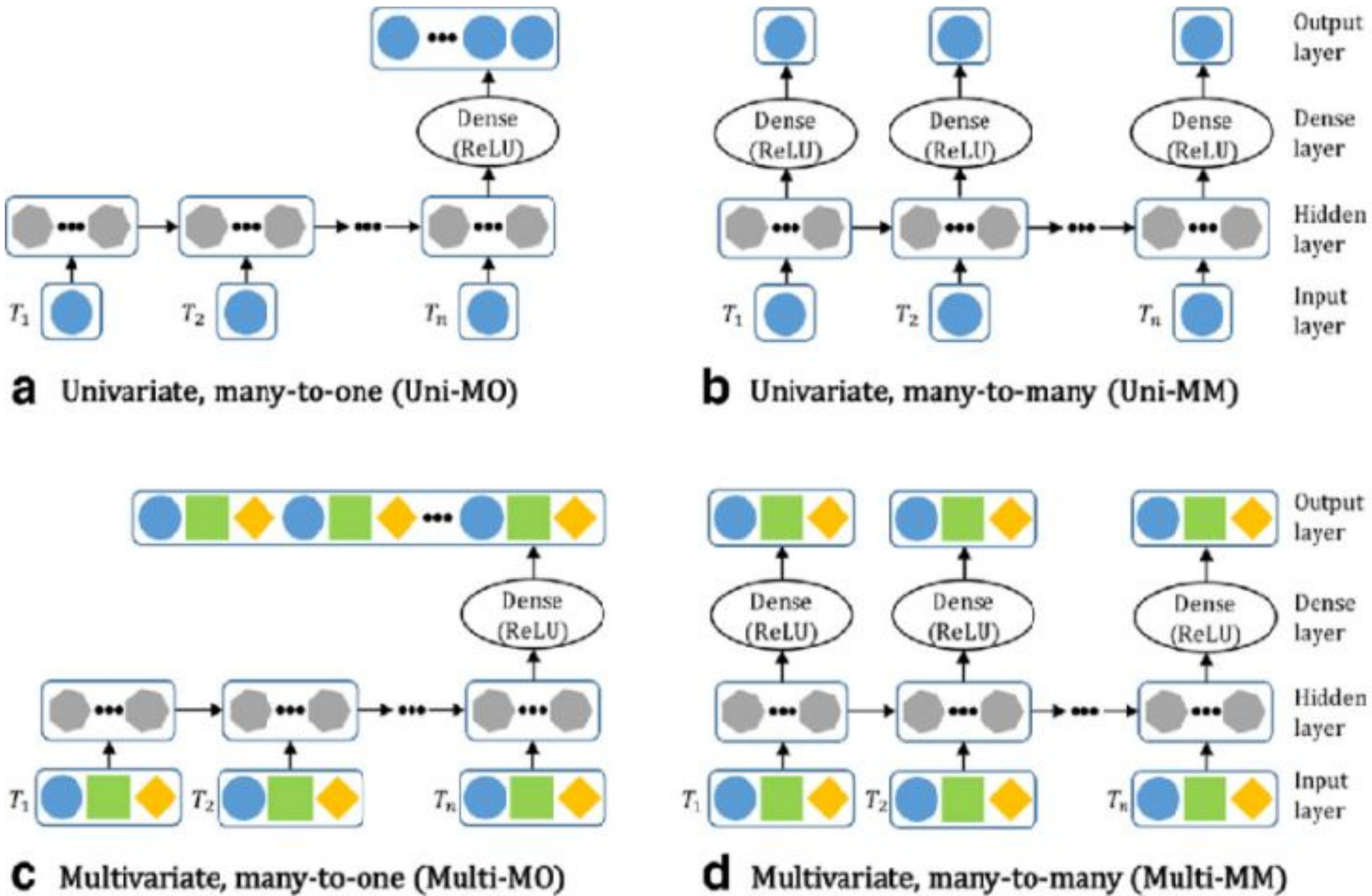
Exercise #8

- Building the RNN LSTM model using sequential model
  - 4 LSTM layers with 50 units (return_sequences=True except for the last layer)
  - Dropout right after LSTM layers with .2
  - Dense with 1 units
  - Compile with adam optimizer and mean_squared_error loss
  - Fit with 100 epochs and 32 batch size

Exercise #8

- Predict with x_train and x_test and inverse transform
- Calculate the root mean squared error
- Visualize the result (vstack and plt.plot)

Types of RNN


Output layer
Dense layer
Hidden layer
Input layer
Dense (ReLU)
T_1
T_2
T_n
a Univariate, many-to-one (Uni-MO)
b Univariate, many-to-many (Uni-MM)
c Multivariate, many-to-one (Multi-MO)
d Multivariate, many-to-many (Multi-MM)

Univariate LSTM Models

- Vanilla LSTM
- Stacked LSTM
- Bidirectional LSTM
- CNN LSTM
- ConvLSTM

```
[10 20 30] 40
[20 30 40] 50
[30 40 50] 60
[40 50 60] 70
[50 60 70] 80
[60 70 80] 90
```

Vanilla LSTM

- An LSTM model that has a single hidden layer of LSTM units, and an output layer used to make a prediction. 히든레이어가 한개이고 출력레이어가 한개인 모델

```
X =
X.reshape((X.shape[0],
X.shape[1], 1))
```

```
model = Sequential()

model.add(LSTM(50,
activation='relu',
input_shape=(3, 1)))

model.add(Dense(1))

model.compile(optimizer=
'adam', loss='mse')

model.fit(X,y,
epochs=200, verbose=0)
```

Stacked LSTM

- Multiple hidden LSTM layers are stacked one on top of another 여러개의 히든레이어가 쌓여있음

```
X = X.reshape((X.shape[0], X.shape[1], 1))
```

```
model = Sequential()

model.add(LSTM(50, activation='relu', return_sequences=True, input_shape=(3, 1)))

model.add(LSTM(50, activation='relu'))

model.add(Dense(1))

model.compile(optimizer='adam', loss='mse')

model.fit(X,y, epochs=200, verbose=0)
```

Bidirectional LSTM

- Allows the LSTM model to learn the input sequence both forward and backwards and concatenate both interpretations. 입력값을 앞뒤로 읽은 결과를 합쳐서 예측

```
X =
X.reshape((X.shape[0],
X.shape[1], 1))
```

```
model = Sequential()
model.add(Bidirectional(LSTM
(50, activation='relu'),
input_shape=(3, 1)))
model.add(Dense(1))
model.compile(optimizer='ada
m', loss='mse')
model.fit(X,y, epochs=200,
verbose=0)
```

CNN LSTM

- A CNN model is used to interpret subsequences of input that together are provided as a sequence to an LSTM model to interpret. **입력값을 나눠서 CNN으로 처리후 합쳐서 LSTM에 입력값으로 줌**

- A number of time steps = 4

```
[10 20 30 40] 50
[20 30 40 50] 60
[30 40 50 60] 70
[40 50 60 70] 80
[50 60 70 80] 90
[60 70 80 90] 100
```

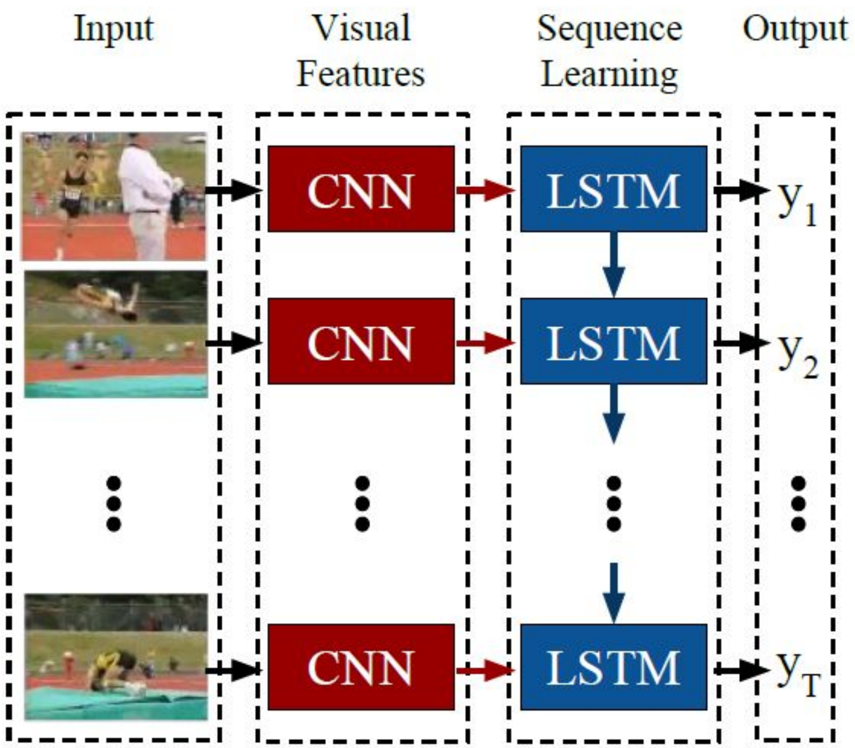
CNN LSTM Code

- Reshape into [samples, subsequences, timesteps, features]
  - n_seq = 2
  - n_steps = 2

```
X = X.reshape((X.shape[0], 2, 2, 1))
```

```
model.add(TimeDistributed(Conv1D(filters=64, kernel_size=1, activation='relu'), input_shape=(None, 2, 1)))

model.add(TimeDistributed(MaxPooling1D(pool_size=2)))

model.add(TimeDistributed(Flatten()))

model.add(LSTM(50, activation='relu'))

model.add(Dense(1))

model.fit(X,y, epochs=500, verbose=0)
```

CNN LSTM Model



- The Time Distributed wrapper allows us to apply a layer to every temporal slice of an input. 임시적으로 만들어지는 입력값에 시간분배 랩퍼 레이어를 적용할수 있음

Exercise #8

https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

- Use the shampoo data to build a vanilla LSTM, stacked LSTM, bidirectional LSTM, and CNN LSTM model and compare the results
- Use the pollution data (pm2.5 column) to build a vanilla LSTM, stacked LSTM, bidirectional LSTM, and CNN LSTM model and compare the results

Generating Dates

- You can specifying the periods and the frequency to create sequence of dates 기간과 빈도를 정해서 데이트레이지를 만듬
- By default, the frequency of range is Days. 일자가 기본값임

```
Import pandas as pd
pd.date_range('1/1/2011', periods=5)
pd.date_range('1/1/2011', periods=5,freq='M')
```

Time Series Frequencies

Alias	Description	Alias	Description
B	business day frequency	BQS	business quarter start frequency
D	calendar day frequency	A	annual(Year) end frequency
W	weekly frequency	BA	business year end frequency
M	month end frequency	BAS	business year start frequency
SM	semi-month end frequency	BH	business hour frequency
BM	business month end frequency	H	hourly frequency

Alias	Description	Alias	Description
MS	month start frequency	T, min	minutely frequency
SMS	SMS semi month start frequency	S	secondly frequency
BMS	business month start frequency	L, ms	milliseconds
Q	quarter end frequency	U, us	microseconds
BQ	business quarter end frequency	N	nanoseconds
QS	quarter start frequency		

Umbrella Sales Example

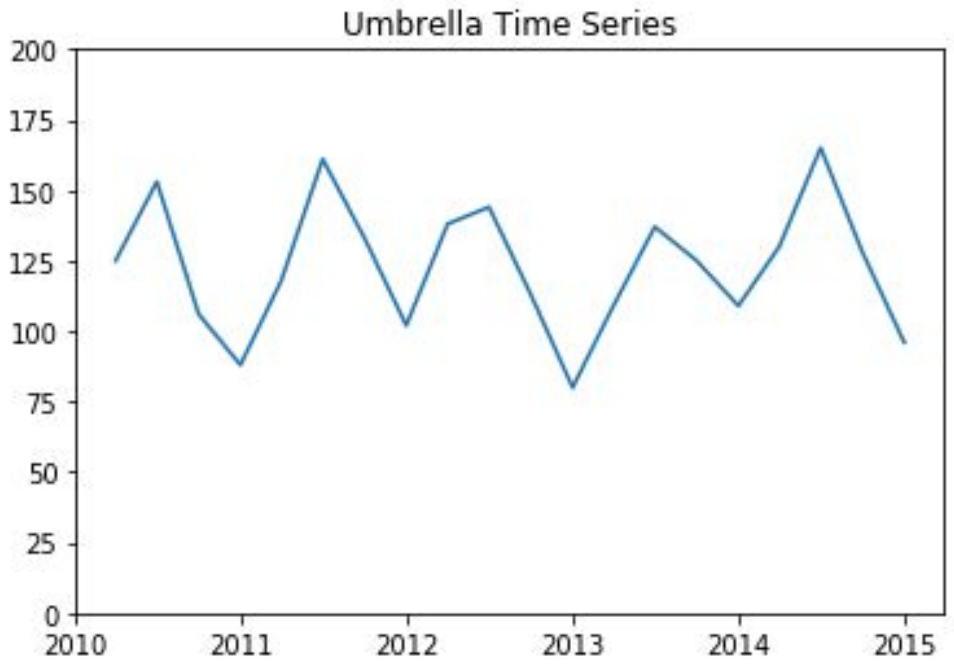
Year	Quarter	Sales
2010	1	125
	2	153
	3	106
	4	88
2011	1	118
	2	161
	3	133
	4	102
2012	1	138
	2	144
	3	113
	4	80
2013	1	109
	2	137
	3	125
	4	109
2014	1	130
	2	165
	3	128
	4	96

This table contains the number of umbrella sold at a clothing store over the past five years. Forecast the quarterly sales for 2015.

이 테이블은 지난 5년간 옷가게에서 팔린 우산의 갯수를 보여줍니다. 2015년의 분기별 세일을 예측하시요

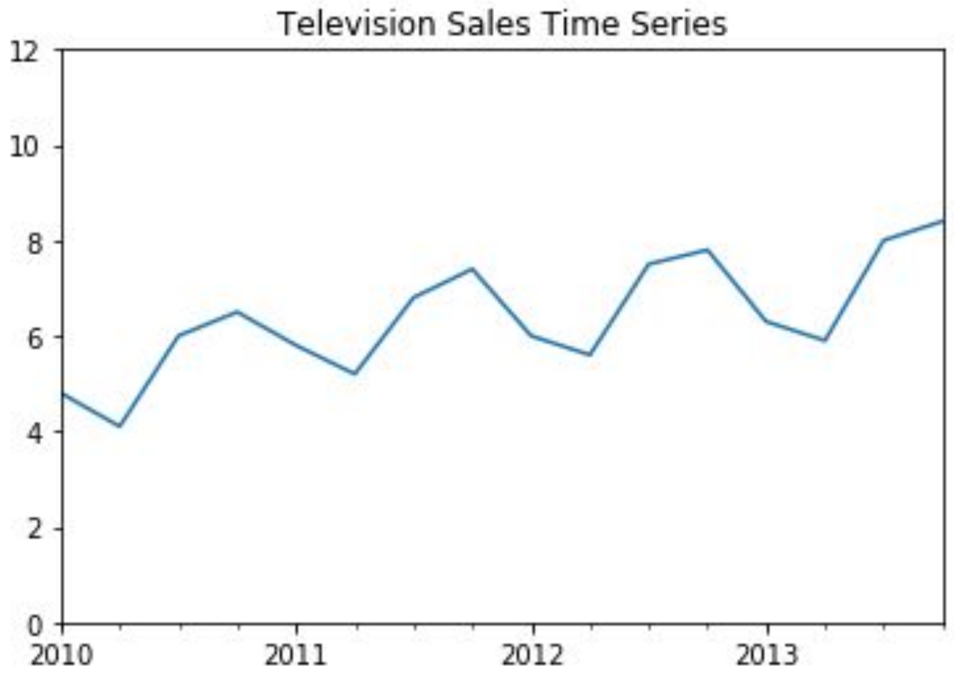
Creating Time Series

```
dt =
pd.date_range('2010-01',
periods=20, freq='Q')

sales =
[125,153,106,88,118,161,13
3,102,138,144,113,80,109,1
37,125,109,130,165,128,96]

df =
pd.DataFrame({'dt':dt,

'sales':sales}, index=dt)

ts = df['sales']

plt.plot(ts)

plt.ylim(0,200)
```



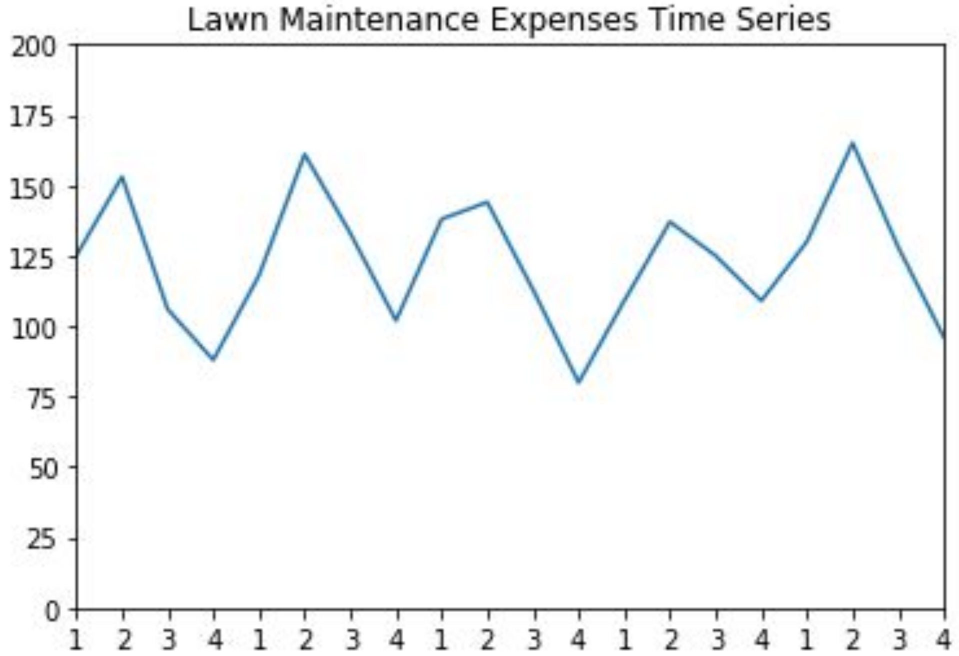
Television Set Sales Example

Year	Quarter	Sales (1000s)
2010	1	4.8
	2	4.1
	3	6
	4	6.5
2011	1	5.8
	2	5.2
	3	6.8
	4	7.4
2012	1	6
	2	5.6
	3	7.5
	4	7.8
2013	1	6.3
	2	5.9
	3	8
	4	8.4



Lawn-Maintenance Expense

Month	2011	2012	2013
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

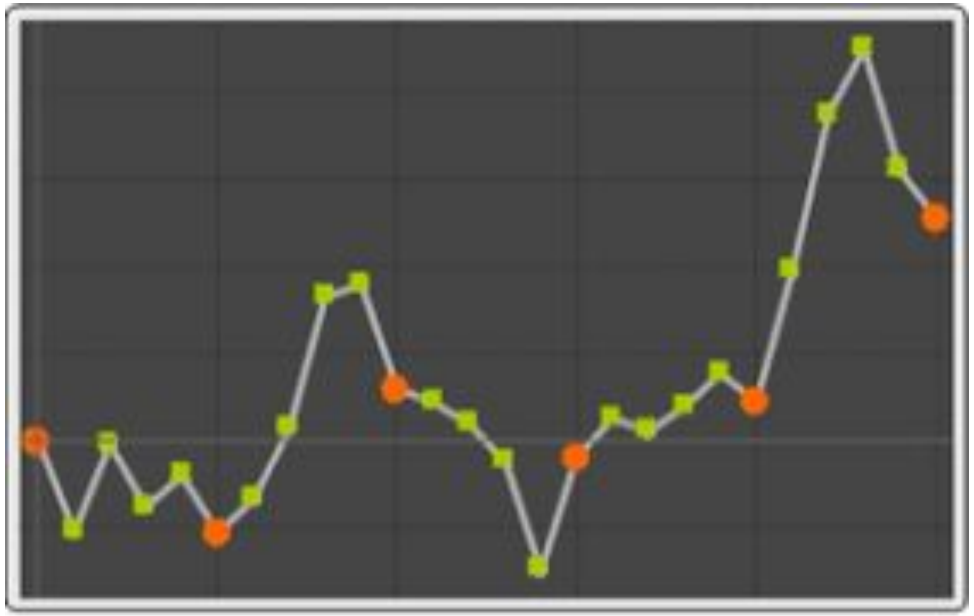


Exercise #8

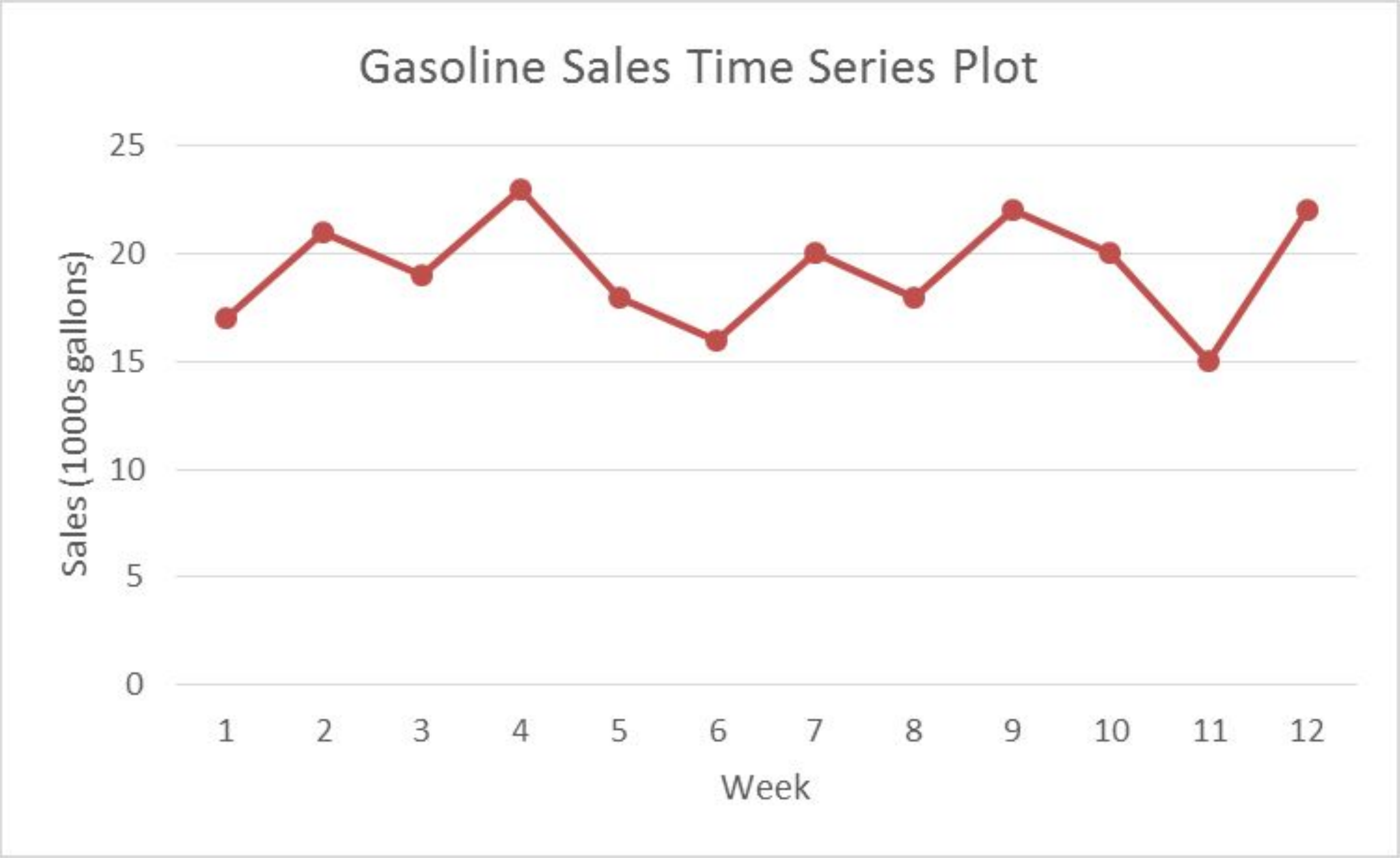
- Umbrella Sales Example
  - Create an umbrella time series (freq=‘Q’) and plot it using plt.plot
- TV Sets Sales Example
  - Create a tvset time series (freq=‘Q’) and plot it using plt.plot
- Lawn-Maintenance Expense Example
  - Create a lawn-maintenance time series (freq=‘M’) and plot it using plt.plot

Time Series (시계열)

- A sequence of observations on a variable measured at successive points in time or over successive periods of time
  시간에 따라 변하는 데이터의 순서

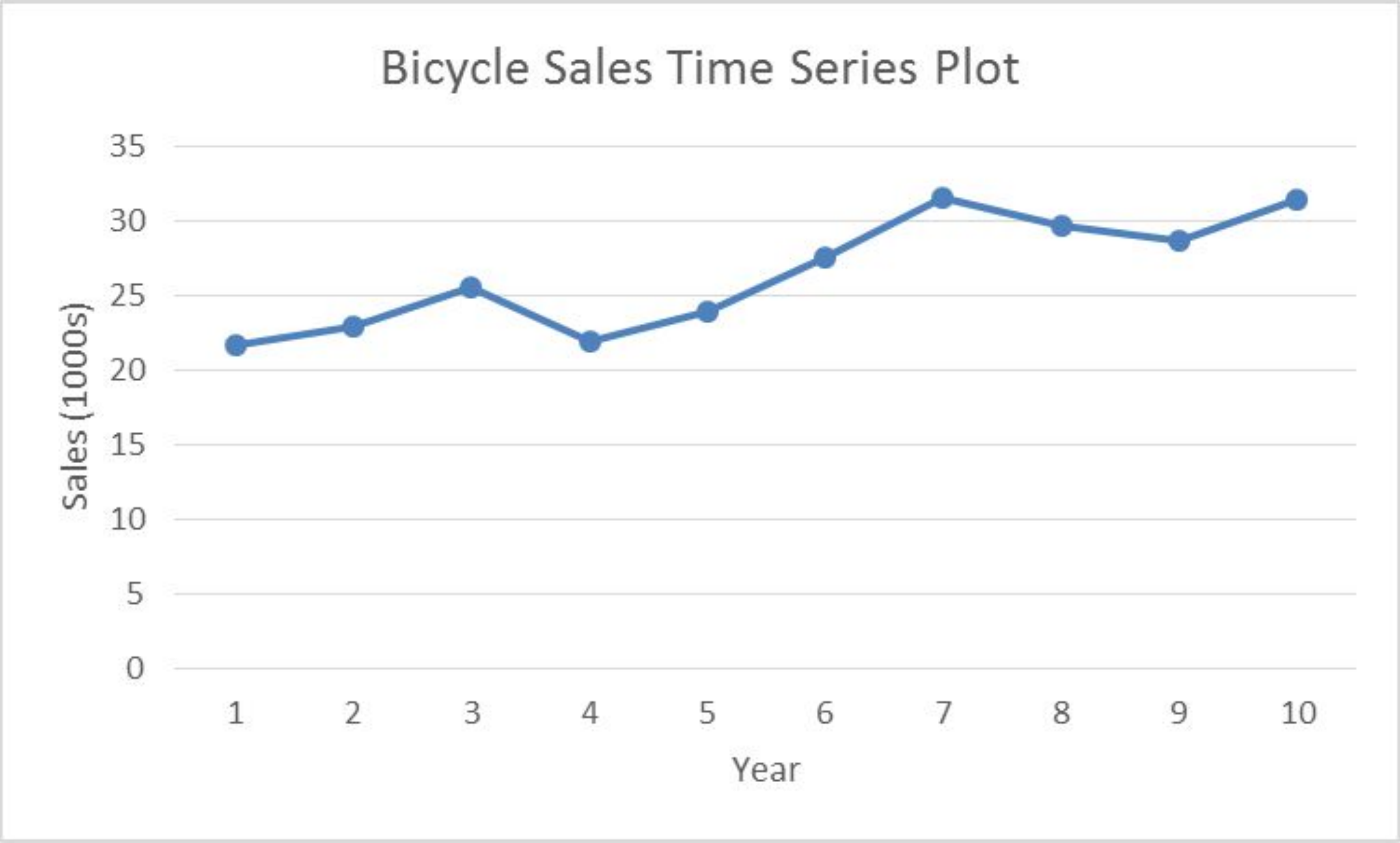


Horizontal Pattern (수평패턴)



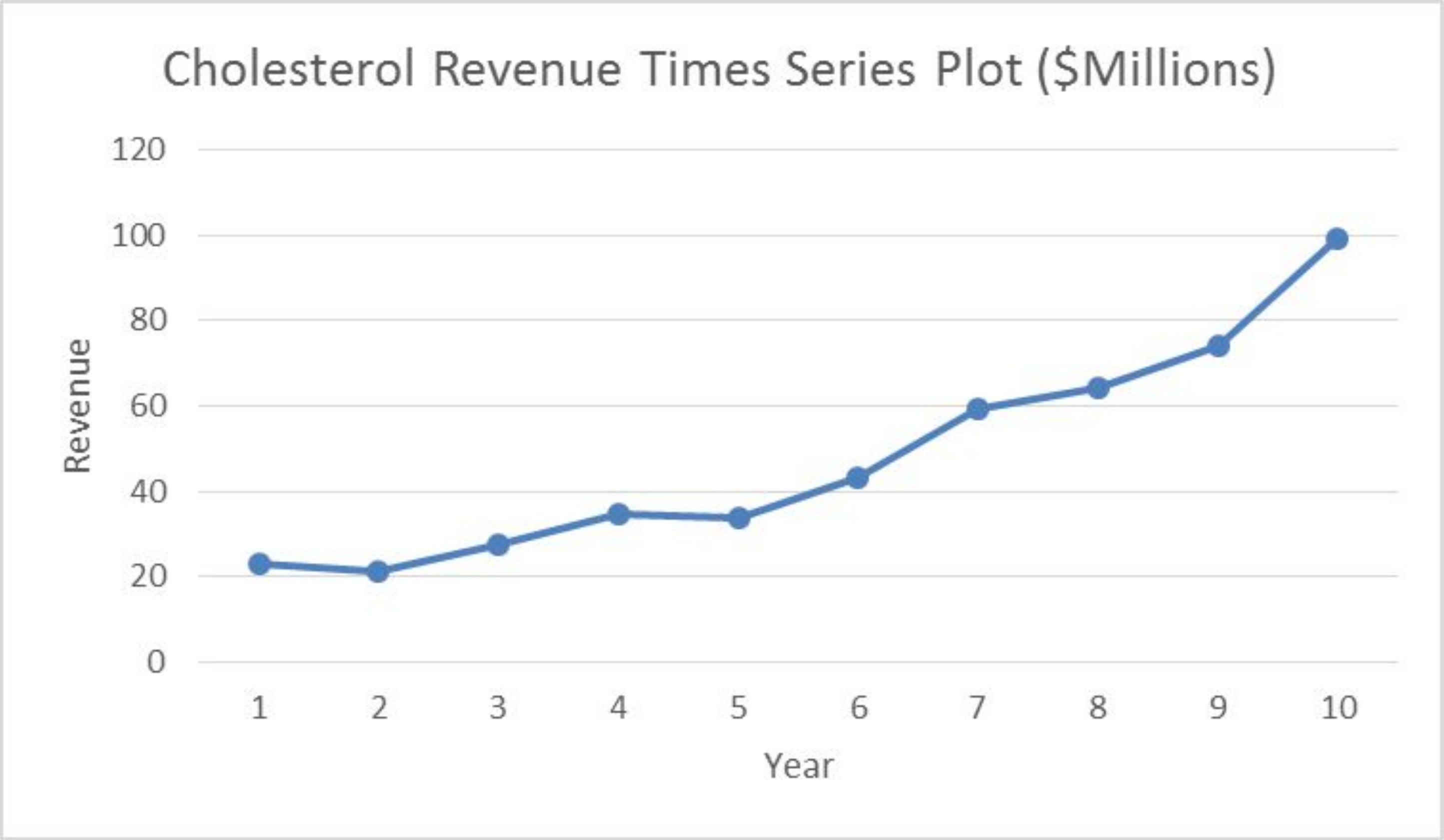
These data follow a horizontal pattern.

Trend Pattern (상향추세)



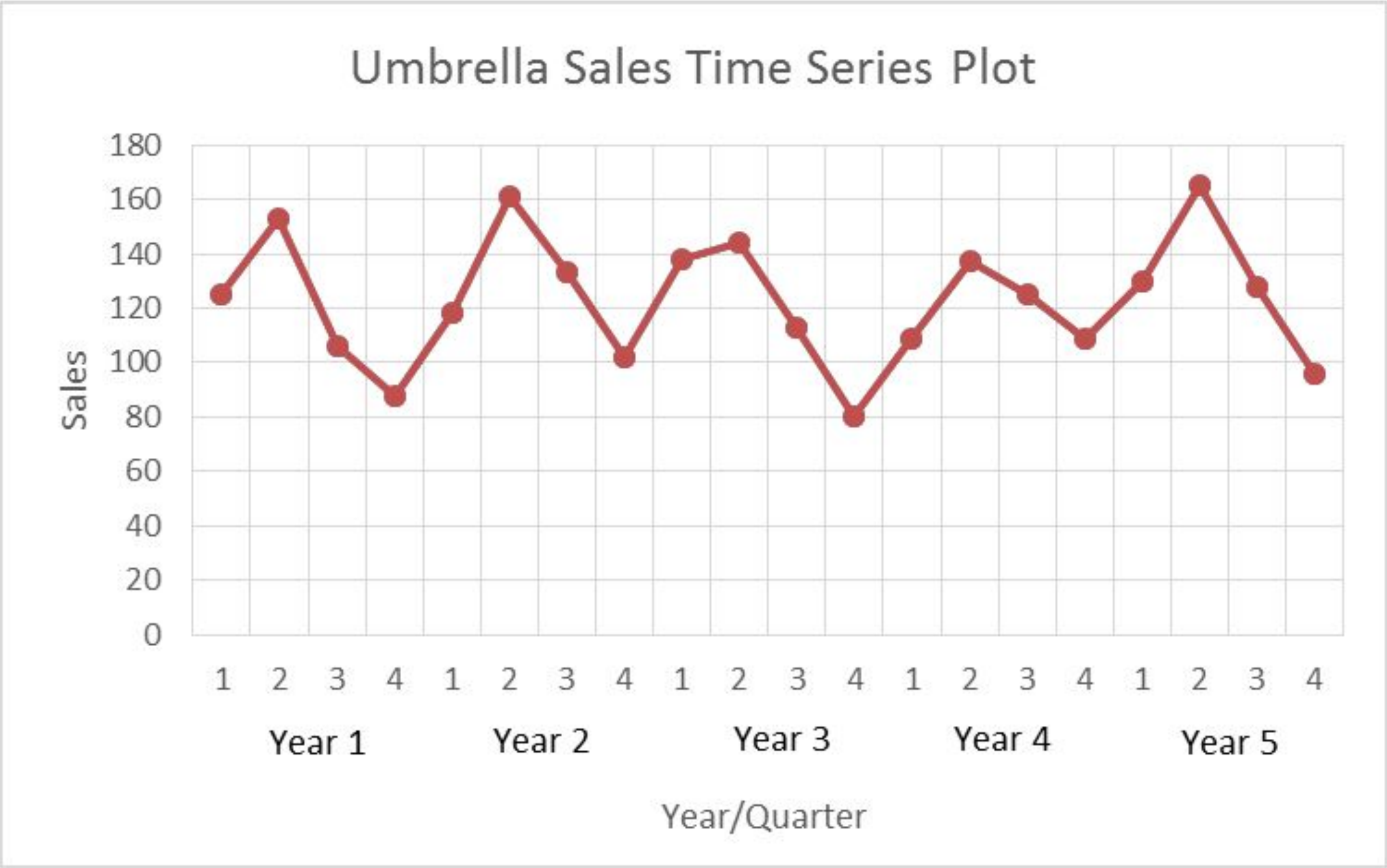
These data have an increasing or upward trend.

Trend Pattern (지수증가패턴)



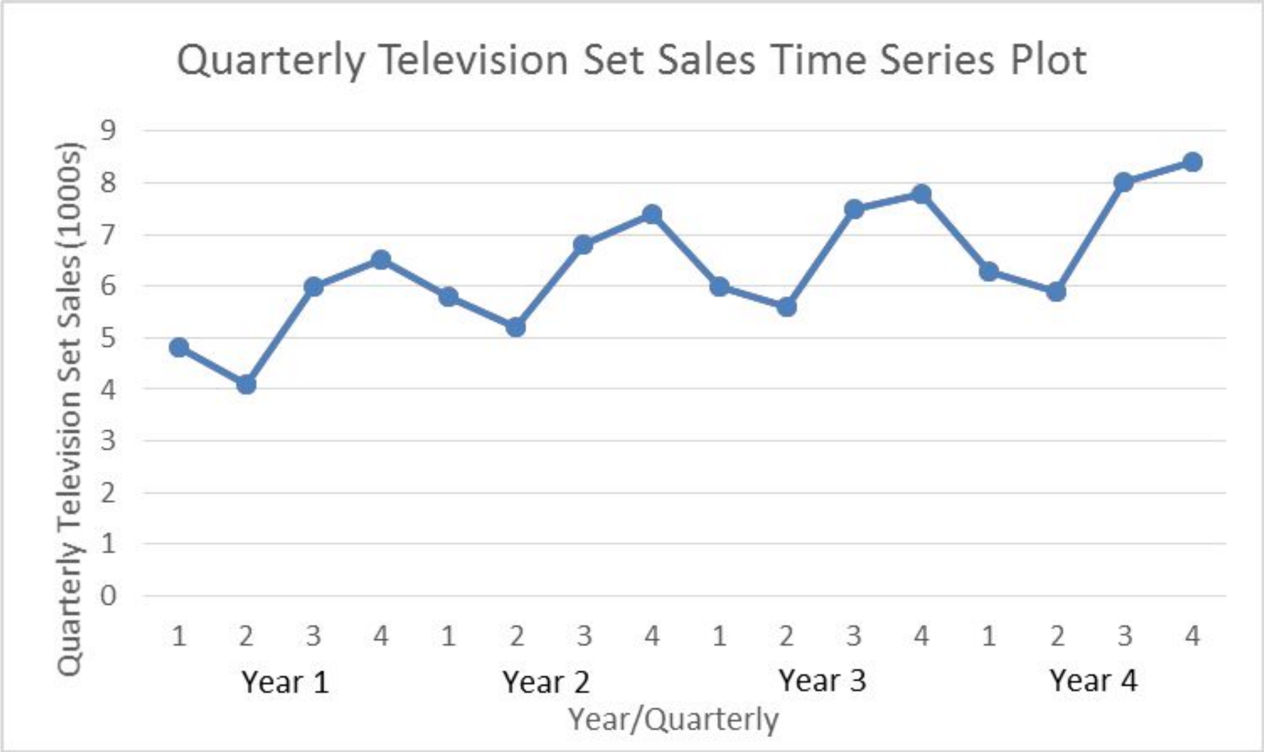
The revenue appears to be growing in an exponential fashion.

Seasonal Pattern (4분기 계절패턴)



A quarterly seasonal pattern is present.

Seasonal Pattern (상향 & 계절패턴)



An increasing trend is present and a seasonal pattern also exists for TV set sales.

Selecting Appropriate Forecasting Techniques (예측기법의 선택)

Forecasting Techniques	Amount of Data	Data Pattern	Forecast Horizon
Trend Regression	10 to 20 observations	Trend (upward or downward)	3 months to 2 years
Trend + Seasonality	At least 5 observation per season	Trend and seasonality	3 months to 2 years
MA	6 to 12 months	No trend or seasonality	Under 3 months
WMA	5 to 10 observations	No trend or seasonality	Under 3 months
Exp. Smoothing	5 to 10 observations	No trend or seasonality	Under 3 months

Summary of Forecasting Methods (예측법 요약)

No Trend □ MA, WMA, exponential smoothing, HoltWinters (simple)

Trend □ Linear regression, quadratic, exponential equation, HoltWinters (double)

Trend + additive seasonal □ dummy, decomposition, HoltWinters (triple)

Time Series Models (시계열모델의 종류)

Smoothing Models

- Simple Moving Average
- Weighted Moving Average
- Exponential Smoothing

Trend Models

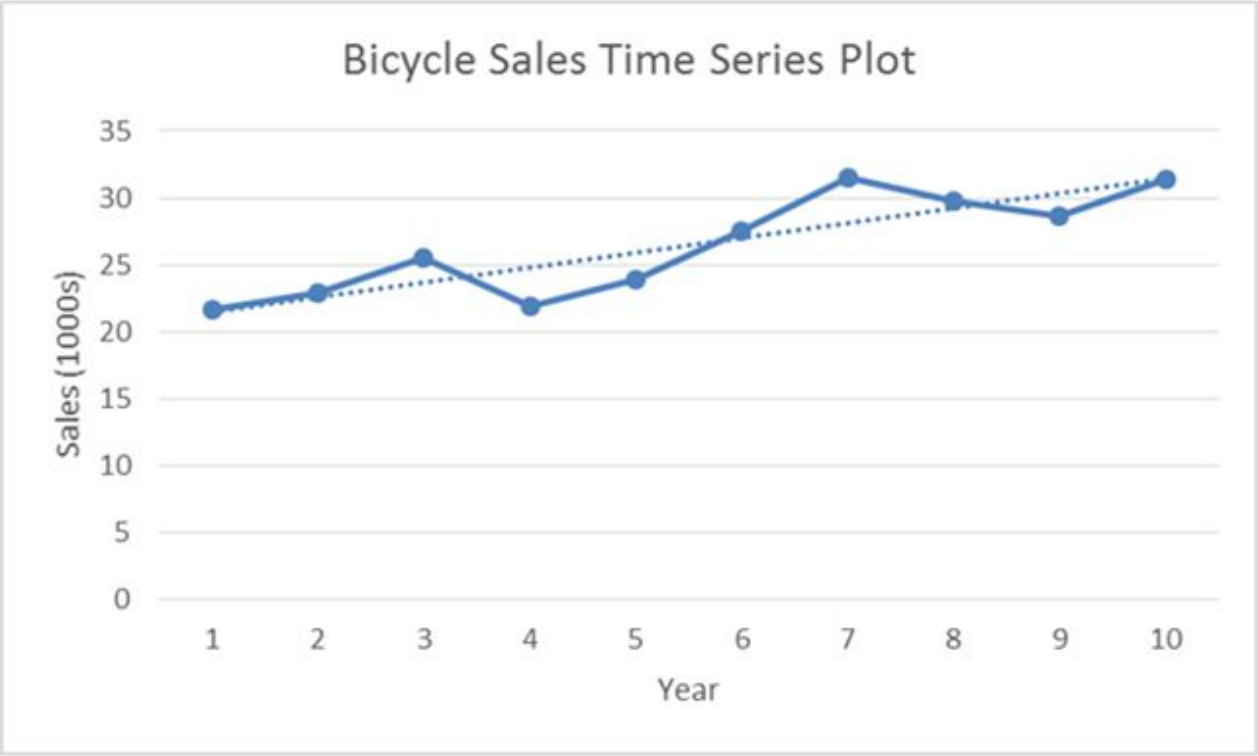
- Linear
- Quadratic
- Exponential
- Auto regression

Linear Trend Regression Example (자전거세일 선형추세 예제)

Year	Sales (1000s)
1	21.6
2	22.9
3	25.5
4	21.9
5	23.9
6	27.5
7	31.5
8	29.7
9	28.6
10	31.4
11	?
12	?

- This table shows the time series of bicycle sales for a particular manufacturer over the past 10 years. Forecast the sales of bicycles for year 11 and 12.

Line Chart (선형차트)



These data have an increasing or upward trend. □
Assumes data falls in a straight line.

Linear Trend Regression (선형추세 방정식)

- = linear trend forecast in period t
- = intercept of the linear trend line
- = slope of the linear trend line
- t the independent variable

Slope and Intercept for Linear Trend (기울기, Y절편)

- Slope $b_1 = \frac{\sum_{t=1}^{n}(t-\bar{t})(Y_t-\bar{Y})}{\sum_{t=1}^{n}(t-\bar{t})^2}$
- Intercept $b_0 = \bar{Y} - b_1\bar{t}$

Where

Y_t = value of the time series in period t
N = number of time periods (number of observations)
$\bar{Y}$ = average value of the time series
$\bar{t}$ = average value of t

Regression Analysis (회귀분석)

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.874526167							
R Square	0.764796016							
Adjusted R Square	0.735395518							
Standard Error	1.958953802							
Observations	10							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	99.825	99.825	26.01302932	0.000929509			
Residual	8	30.7	3.8375					
Total	9	130.525						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	20.4	1.338220211	15.24412786	3.39989E-07	17.31405866	23.48594134	17.31405866	23.48594134
Year	1.1	0.215673715	5.100296983	0.000929509	0.60265552	1.59734448	0.60265552	1.59734448

Linear Trend Equation

$$T_t = 20.4 + 1.1t$$

Forecast Using Linear Equation (선형방정식을 통한 예측)

Slope	1.1	
Intercept	20.4	
Year	Sales (1000s)	Forecast
1	21.6	21.5
2	22.9	22.6
3	25.5	23.7
4	21.9	24.8
5	23.9	25.9
6	27.5	27
7	31.5	28.1
8	29.7	29.2
9	28.6	30.3
10	31.4	31.4
11		32.5
12		33.6

x

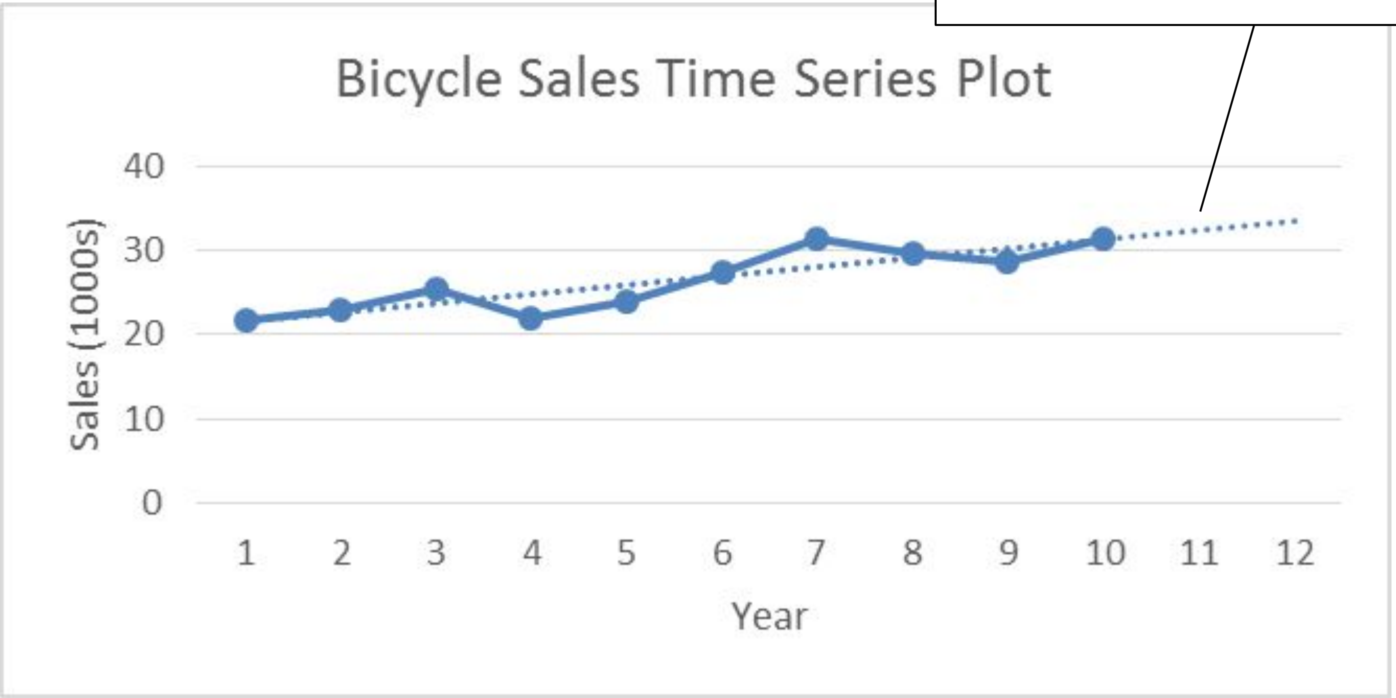
y

$$= 20.4 + 1.1 \times 11$$

$$= 20.4 + 1.1 \times 12$$

Linear Chart (선형차트)

sns.lmplot(x,y, data)

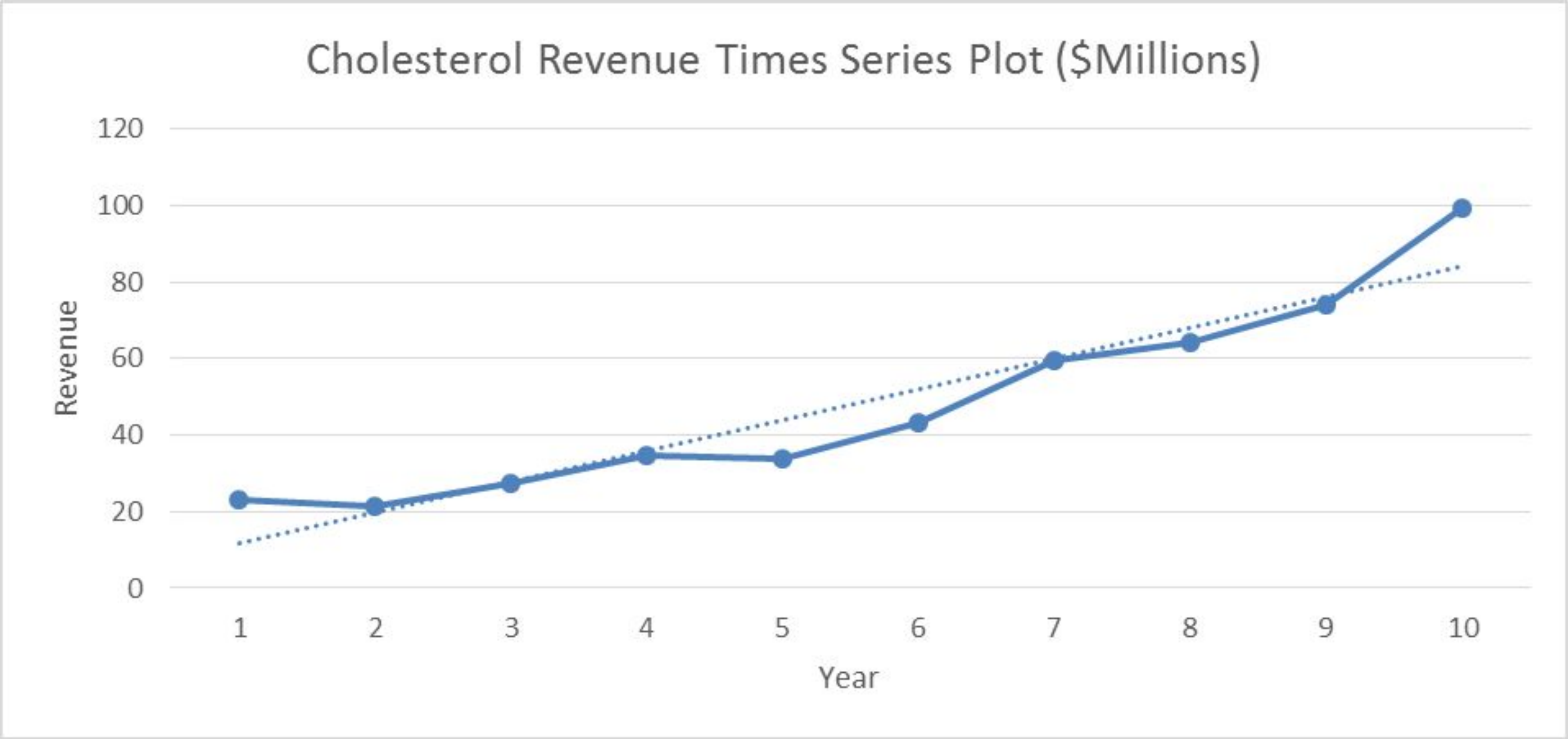


Nonlinear Trend Regression Example (콜레스테롤약 판매 비선형추세 예제)

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	?

- 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.

Scatter Chart (산점도)



The revenue appears to be growing in an exponential fashion.

□ Assumes data falls in quadratic trend line.

Quadratic Trend Equation (이차추세 방정식)

- $$T_t = b_0 + b_1 t + b_2 t^2$$

Year	Sq. Year	Revenue
1	1	23.1
2	4	21.3
3	9	27.4
4	16	34.6
5	25	33.8
6	36	43.2
7	49	59.5
8	64	64.4
9	81	74.2
10	100	99.3

= Year ^2

Regression Analysis

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.990547732
R Square	0.981184809
Adjusted R Square	0.97580904
Standard Error	3.975782479
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	5770.128076	2885.064038	182.5199018	9.13651E-07
Residual	7	110.6479242	15.80684632		
Total	9	5880.776			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	24.18166667	4.676124187	5.171305487	0.001293222	13.12439001	35.23894332	13.12439001	35.23894332
Year	-2.105984848	1.952946599	-1.078362741	0.316623126	-6.723969739	2.512000042	-6.723969739	2.512000042
Sq. Year	0.921590909	0.173023724	5.32638467	0.001091702	0.512454816	1.330727002	0.512454816	1.330727002

Quadratic Trend Equation

$$T_t = 24.1817 + 2.1060\, t + 0.9216\, t^2$$

Forecast Using Quadratic Equation

Year	Sq. Year	Revenue	Forecast
1	1	23.1	23.00
2	4	21.3	23.66
3	9	27.4	26.16
4	16	34.6	30.50
5	25	33.8	36.69
6	36	43.2	44.72
7	49	59.5	54.60
8	64	64.4	66.32
9	81	74.2	79.88
10	100	99.2	95.28
11	121		112.53
12	144		131.62

x

y

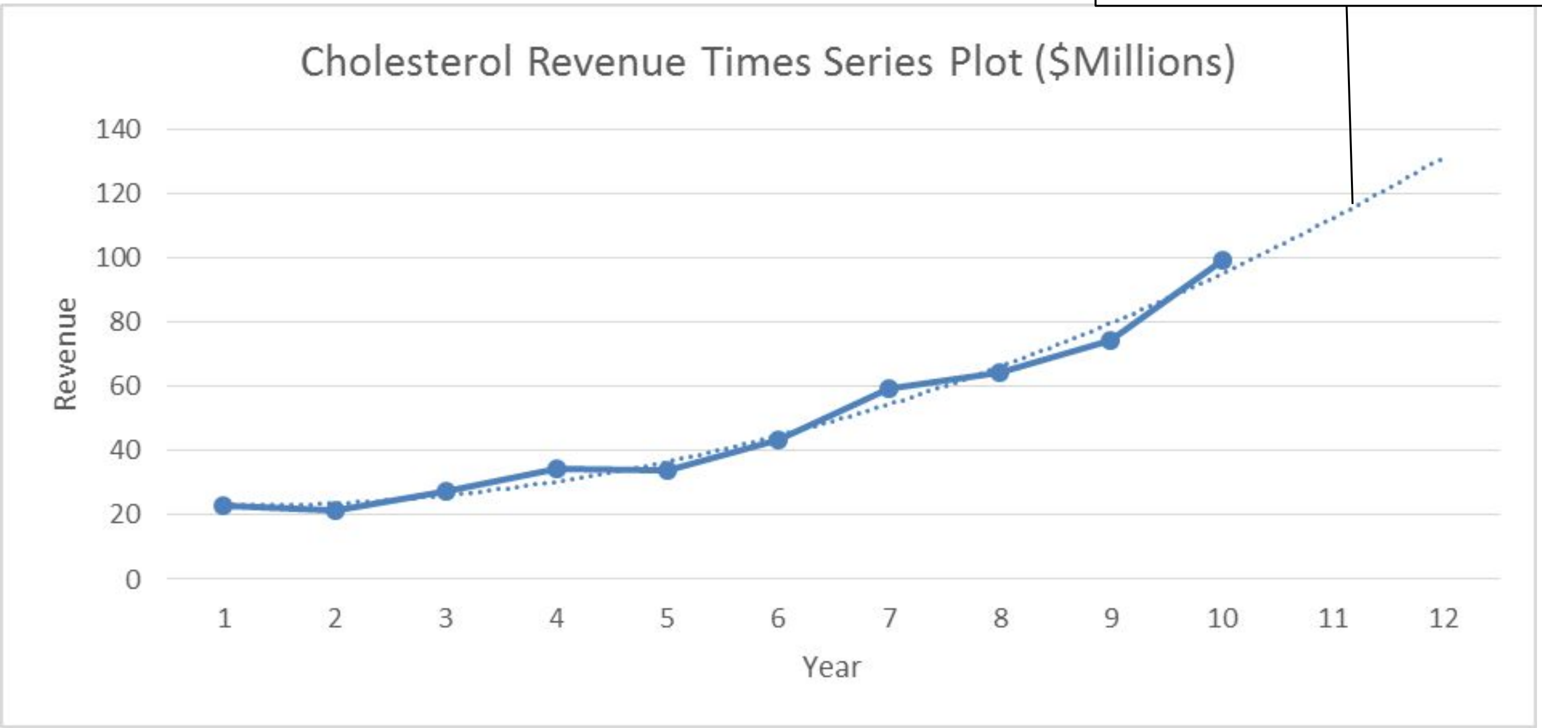
$= 24.1817 + 2.1060 * 11 + 0.9216 * 11\text{^}2$

$= 24.1817 + 2.1060 * 12 + 0.9216 * 12\text{^}2$

= Intercept + coefficient of Year * Year + coefficient of Sq. Year * Sq. Year

Quadratic Chart (이차식 차트)

Forecast for year 11 and 12



Forecast Errors (예측에러)

- Forecast Error
  = Actual Value – Forecast

Negative errors = over-forecasts
Positive errors = under-forecasts

- Forecasts are never perfect, so need to know how much we should rely on our chosen forecasting method

Forecast Errors (예측에러 계산법)

			ME	MAE	MAPE	MSE
Slope	1.1					
Intercept	20.4		0.00	1.32	5.07%	3.07

Year	Sales (1000s)	Forecast	Error	Abs Error	% Error	Sq. Error
1	21.6	21.5	0.1	0.1	0.46%	0.01
2	22.9	22.6	0.3	0.3	1.31%	0.09
3	25.5	23.7	1.8	1.8	7.06%	3.24
4	21.9	24.8	-2.9	2.9	13.24%	8.41
5	23.9	25.9	-2	2	8.37%	4
6	27.5	27	0.5			
7	31.5	28.1	3.4			
8	29.7	29.2	0.5			
9	28.6	30.3	-1.7			
10	31.4	31.4	0			
11		32.5				
12		33.6				

- Error = Sales – Forecast
- ABS Error = abs(Error)
- % Error = ABS Error / Sales
- Sq. Error = Error^2

Measures of Forecast Errors (예측에러 측정치)

Mean Error (ME)	• $ME = 1/n \sum e_t$
Mean Absolute Error (MAE)	• $MAE = 1/n \sum \lvert e_t \rvert$
Mean Absolute Percent Error (MAPE)	• MAPE = MAD / Actual Sales
Mean Squared Error (MSE)	• $MSE = 1/n \sum e^2_t$

Measures of Forecasting Error

- ME = The arithmetic average of all prediction errors
- MAE = The average of the absolute value of the errors
- MAPE = The absolute error as a percentage of the sales.
- MSE = the arithmetic mean of the sum of the squares of the prediction errors
- TS = The arithmetic sum of forecast deviations divided by MAD.
- RSFE = Runtime Sum of Forecast Errors

Forecast Errors

Slope	1.1		ME	MAE	MAPE	MSE
Intercept	20.4		0.00	1.32	5.07%	3.07

Year	Sales (1000s)	Forecast	Error	Abs Error	% Error	Sq. Error
1	21.6	21.5	0.1	0.1	0.46%	0.01
2	22.9	22.6	0.3	0.3	1.31%	0.09
3	25.5	23.7	1.8	1.8	7.06%	3.24
4	21.9	24.8	-2.9	2.9	13.24%	8.41
5	23.9	25.9	-2	2	8.37%	4
6	27.5	27	0.5	0.5	1.82%	0.25
7	31.5	28.1	3.4	3.4	10.79%	11.56
8	29.7	29.2	0.5	0.5	1.68%	0.25
9	28.6	30.3	-1.7	1.7	5.94%	2.89
10	31.4	31.4	0	0	0.00%	0
11		32.5				
12		33.6				

- ME = mean(Errors)
- MAE = mean(ABS Errors)
- MAPE = mean(% Errors)
- MSE = mean(Sq. Errors)

Tracking Signal (추적신호)

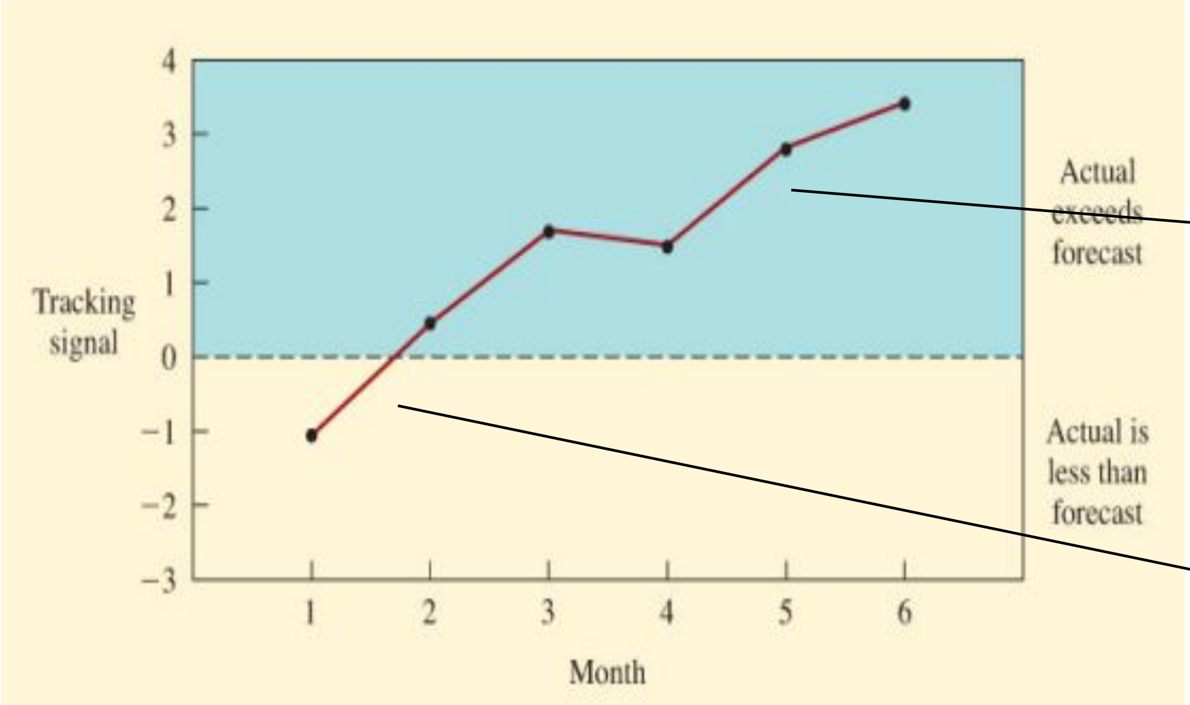
- A measure that indicates whether the forecast average is keeping pace with any genuine upward or downward changes in demand 예측평균이 상승되는지 감소되는 보여주는 측정치

Tracking Signals

Year	Sales (1000s)	Forecast	Error	Abs Error	% Error	Sq. Error	Sum of Error	Sum of AE	MAD	TS
1	21.6	21.5	0.1	0.1	0.46%	0.01	0.10	0.10	0.10	1.00
2	22.9	22.6	0.3	0.3	1.31%	0.09	0.40	0.40	0.20	2.00
3	25.5	23.7	1.8	1.8	7.06%	3.24	2.20	2.20	0.73	3.00
4	21.9	24.8	-2.9	2.9	13.24%	8.41	-0.70	5.10	1.28	-0.55
5	23.9	25.9	-2	2	8.37%	4	-2.70	7.10	1.42	-1.90
6	27.5	27	0.5	0.5	1.82%	0.25	-2.20	7.60	1.27	-1.74
7	31.5	28.1	3.4	3.4	10.79%	11.56	1.20	11.00	1.57	0.76
8	29.7	29.2	0.5	0.5	1.68%	0.25	1.70	11.50	1.44	1.18
9	28.6	30.3	-1.7	1.7	5.94%	2.89	0.00	13.20	1.47	0.00
10	31.4	31.4	0	0	0.00%	0	0.00	13.20	1.32	0.00

- Sum of Errors = cumsum(Errors)
- Sum of AE = cumsum(ABS Errors)
- MAD = Sum of AE / Year
- TS = Sum of Errors / MAD

Tracking Signal

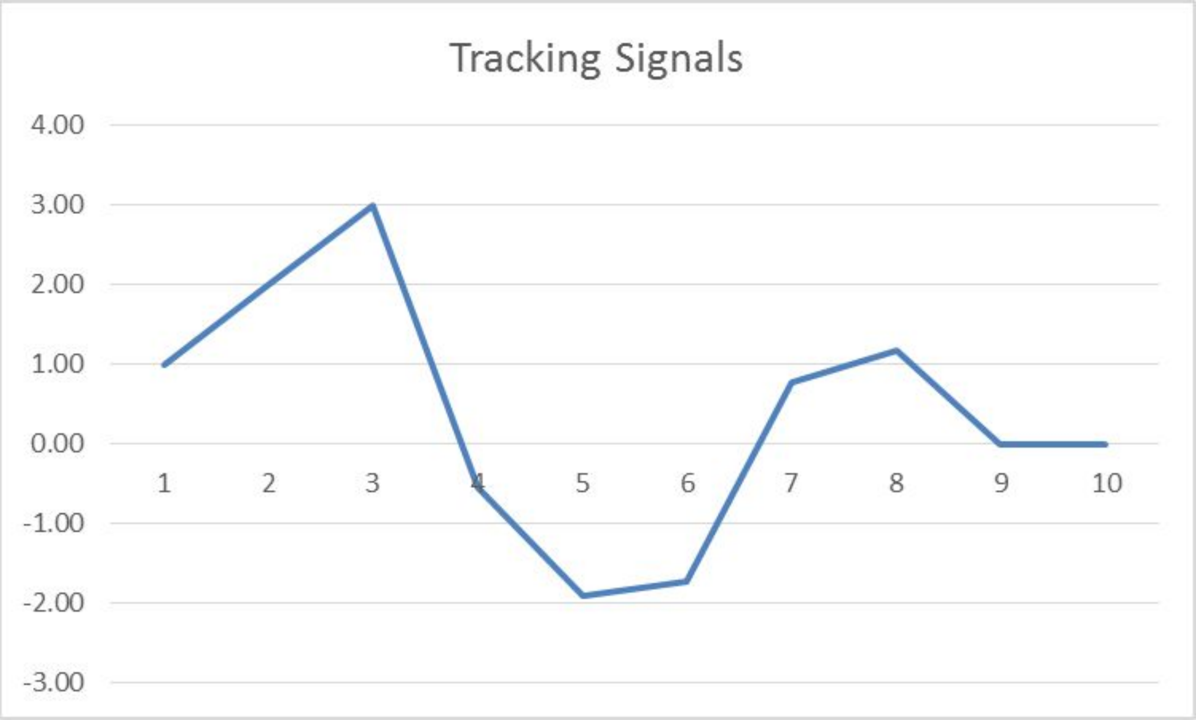


Positive TS - actual exceeds forecast

Negative TS - actual is less than forecast

$$TS = \frac{Running\ Sum\ of\ Forecast\ Errors}{Mean\ Absolute\ Error}$$

Tracking Signal Chart (추적신호차트)



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

Forecast Errors Comparison
예측오류비교

3PMA is the better forecast on ME.

Exponential Smoothing is the better forecast on MAE, MAPE, and MSE.

							ME	MAE	MAPE	MSE	ME	MAE	MAPE	MSE	ME	MAE	MAPE	MSE
							0.00	2.67	14.36%	10.22	0.06	2.98	15.99%	11.49	0.99	2.60	13.40%	8.98
		Forecast					3PMA				3PWMA				Exp			
Week	Sales (1000s of gallons)	3PMA	Wt.	3PWMA	alpha	Exp	Error	ABS Error	% Error	Sq. Error	Error	ABS Error	% Error	Sq. Error	Error	ABS Error	% Error	Sq. Error
1	17		0.166666667		0.2													
2	21		0.333333333			17.00									4.00	4.00	19.05%	16.00
3	19		0.5			17.80									1.20	1.20	6.32%	1.44
4	23	19		19.33		18.04	4.00	4.00	17.39%	16.00	3.67	3.67	15.94%	13.44	4.96	4.96	21.57%	24.60
5	18	21		21.33		19.03	-3.00	3.00	16.67%	9.00	-3.33	3.33	18.52%	11.11	-1.03	1.03	5.73%	1.07
6	16	20		19.83		18.83	-4.00	4.00	25.00%	16.00	-3.83	3.83	23.96%	14.69	-2.83	2.83	17.66%	7.98
7	20	19		17.83		18.26	1.00	1.00	5.00%	1.00	2.17	2.17	10.83%	4.69	1.74	1.74	8.70%	3.03
8	18	18		18.33		18.61	0.00	0.00	0.00%	0.00	-0.33	0.33	1.85%	0.11	-0.61	0.61	3.38%	0.37
9	22	18		18.33		18.49	4.00	4.00	18.18%	16.00	3.67	3.67	16.67%	13.44	3.51	3.51	15.97%	12.34
10	20	20		20.33		19.19	0.00	0.00	0.00%	0.00	-0.33	0.33	1.67%	0.11	0.81	0.81	4.05%	0.66
11	15	20		20.33		19.35	-5.00	5.00	33.33%	25.00	-5.33	5.33	35.56%	28.44	-4.35	4.35	29.01%	18.94
12	22	19		17.83		18.48	3.00	3.00	13.64%	9.00	4.17	4.17	18.94%	17.36	3.52	3.52	15.99%	12.38
13		19		19.33		19.18												

Overall, exponential smoothing is the best forecast because it outperforms the others on 3 out of 4 metrics.

Evaluate Forecast (예측의 평가)

ME	A large positive (negative) ME means that the forecast is undershooting (overshooting) the actual observations 적은게 좋음
MAE	Wants MAD as small as possible 적은게 좋음
MAPE	Most common measure of forecast accuracy; 10% is considered very good; 20% - 30% or even higher is quite common 10%는 좋은편, 20-30%로 괜찮음
TS	As long as TS is between -4 and 4, assume the model is working correctly.

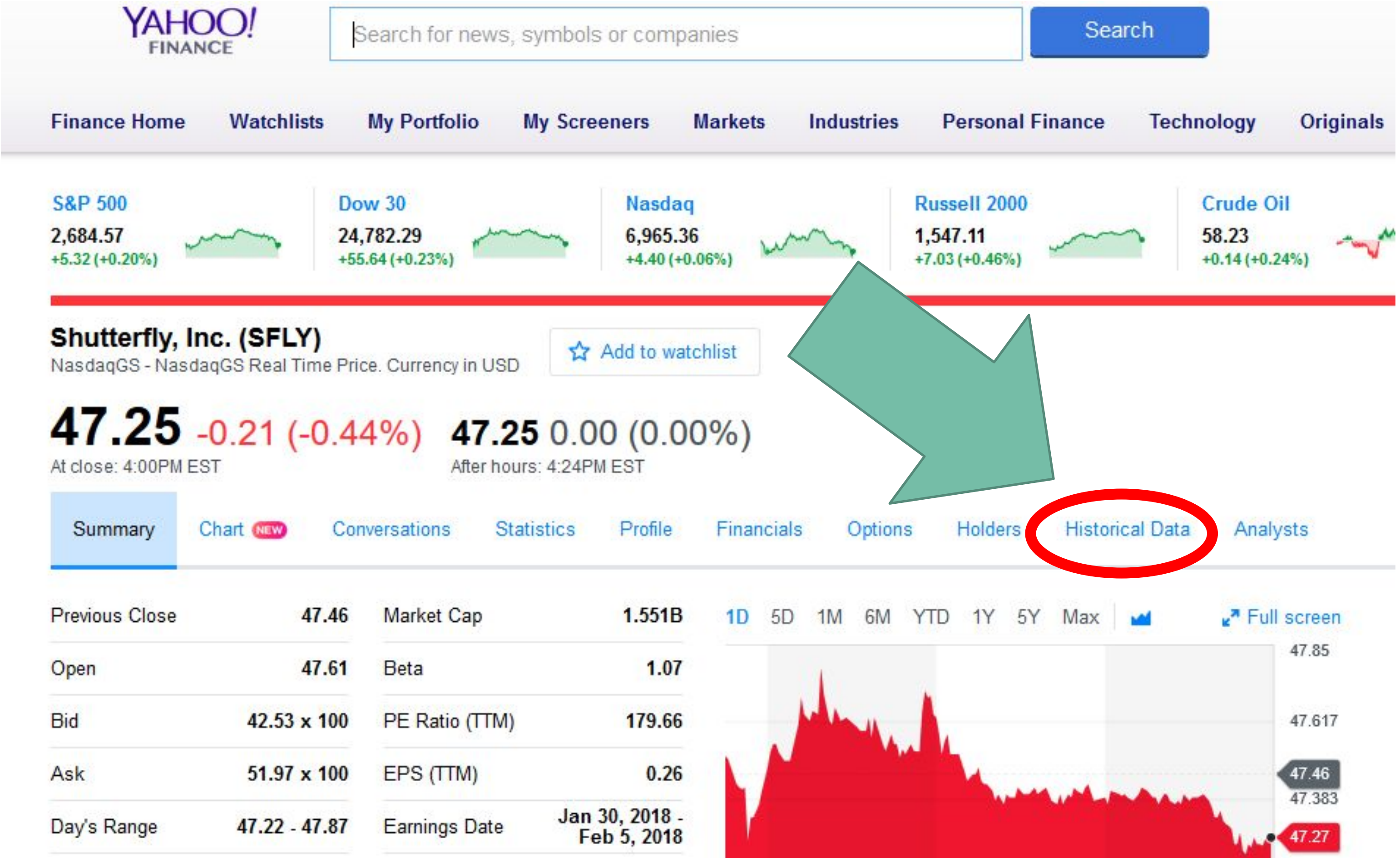
Exercise #8

- Import the bicycle data
  - Plot the time series
  - Conduct a regression analysis and plot the regression equation on the time series plot
  - Forecast year 11 and year 12
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

Exercise #8

- Import the revenue data
  - Plot the time series
  - Conduct a regression analysis and plot the regression equation on the time series plot
  - Forecast year 11 and year 12
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

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
- Use the linear regression to forecast next 30 days’ stock price
- Calculate ME, MAD, MAPE, MSE, and TS
- Draw the forecast chart and tracking signals chart

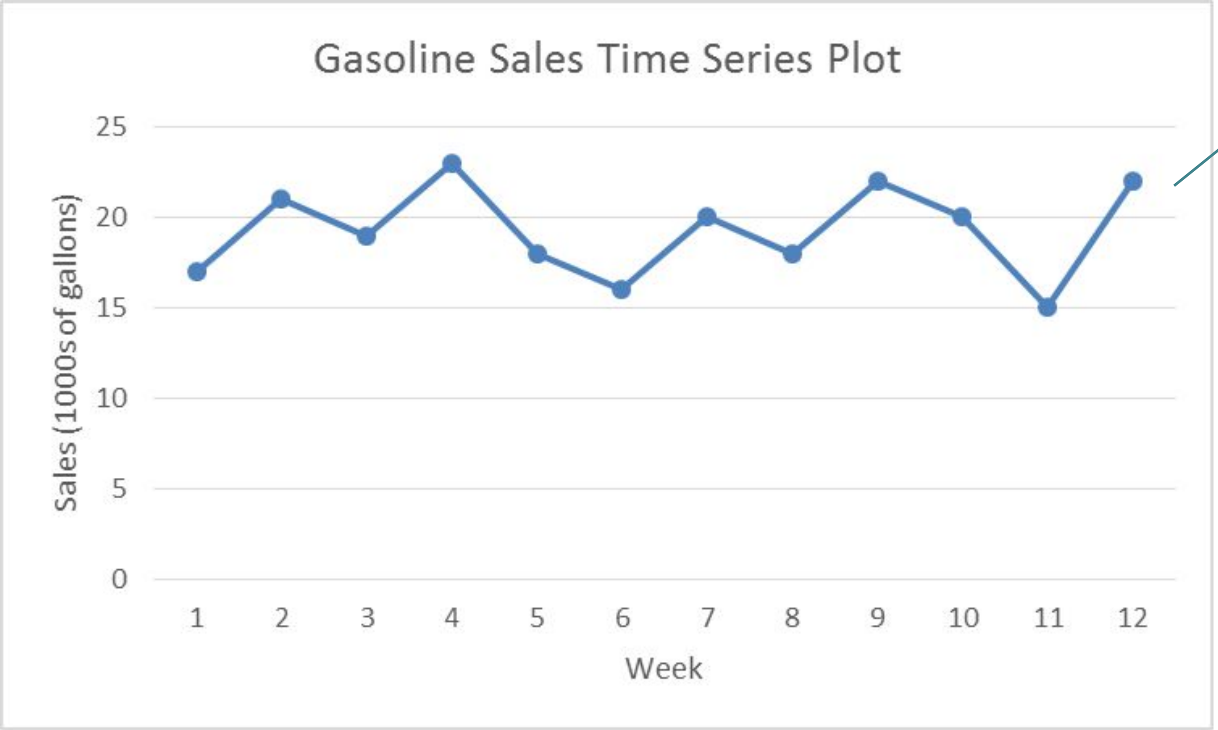
Gasoline Sales Example
가솔린세일 비선형주세 예측

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 예측하십시오.

Time Series Plot (시계열그림)



A horizontal pattern is present!

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

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
- 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.
- Forecast for Week 13
  $= (20 + 18 + 22 + 20 + 15 + 22)/6 = 19.5$

Moving Average Forecast of Order k (k차의 이동평균예측)

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

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

Where

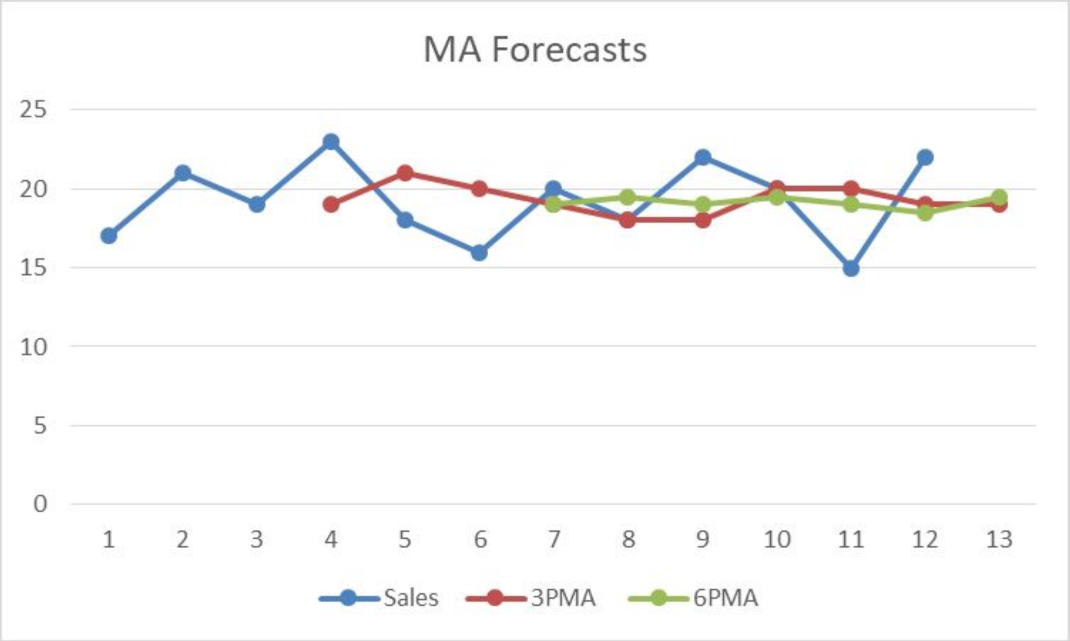
F_{t+1} = forecast of the times series for period t + 1

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.

.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
- Forecast for Week 13 = (3/6)*22 + (2/6)*15 + (1/6)*20 = 19.33

Weighted Moving Average (가중이동평균)

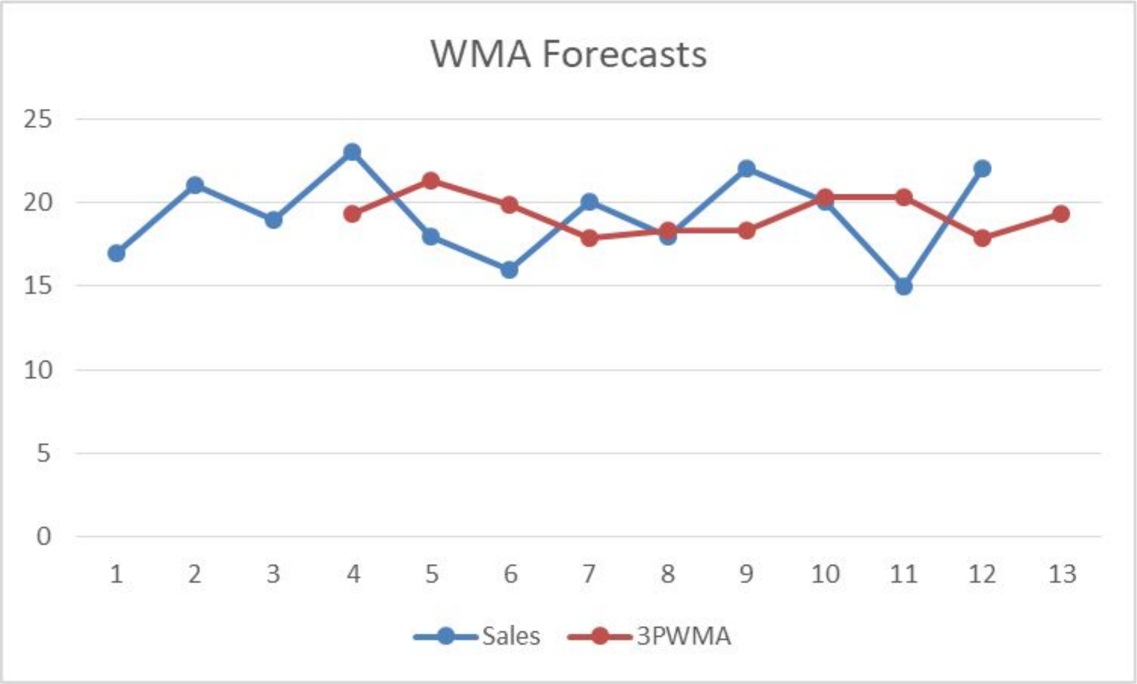
$$F_{t+1} = \sum w_t A_t$$

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 (가중이동평균표)



- 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.
- 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).

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 ($0 \le \alpha \le 1$)

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

Alpha Selection (알파의 선택)

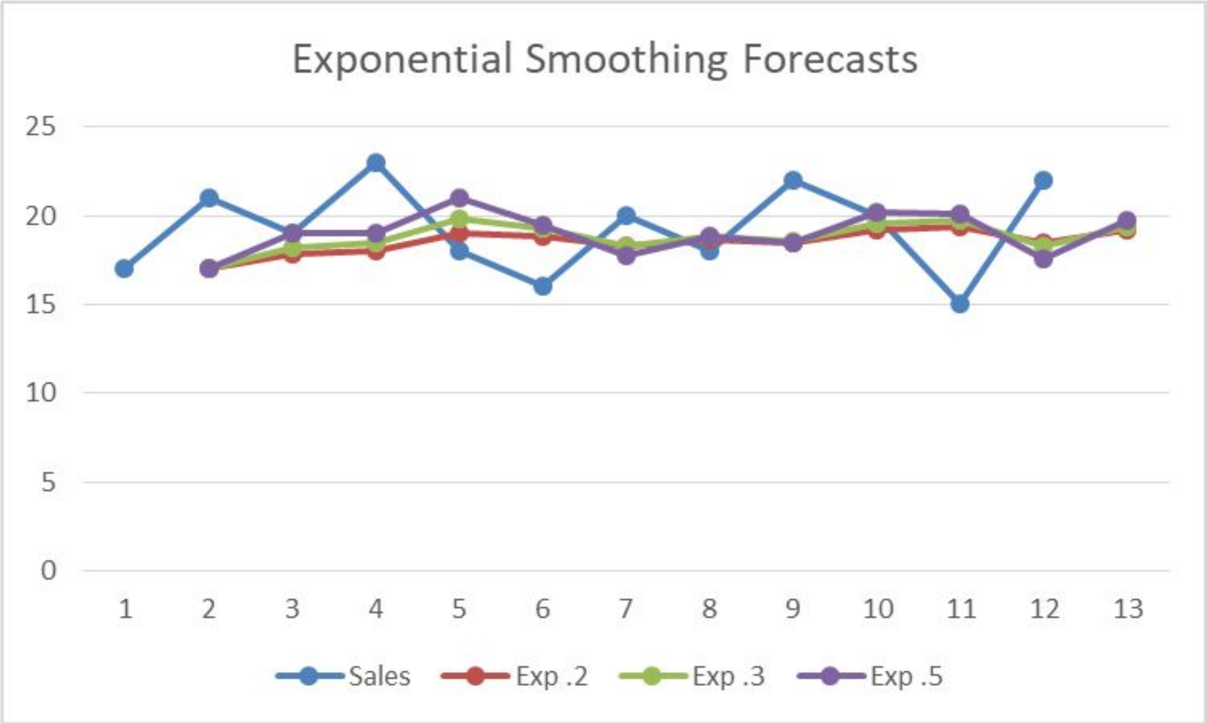
- Use trial and error to determine the value of alpha minimizes the MSE!
  여러번의 시도로 MSE를 최소화하는 알파값을 구함
- If the time series contains substantial random variability, a small value of the smoothing constant is preferred.
  - Larger values of the smoothing constant allows the forecast to react more quickly 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

.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 #8

- Gasoline Sales Example
  - Forecast week 13 using simple calculation, formulas, and forecast functions
  - Simple moving average (3 weeks, 6 weeks)
  - Weighted moving average (3/6, 2/6, 1/6)
  - Exponential smoothing method (alpha = .2, .3, .5)

Exercise #8

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

Seasonality

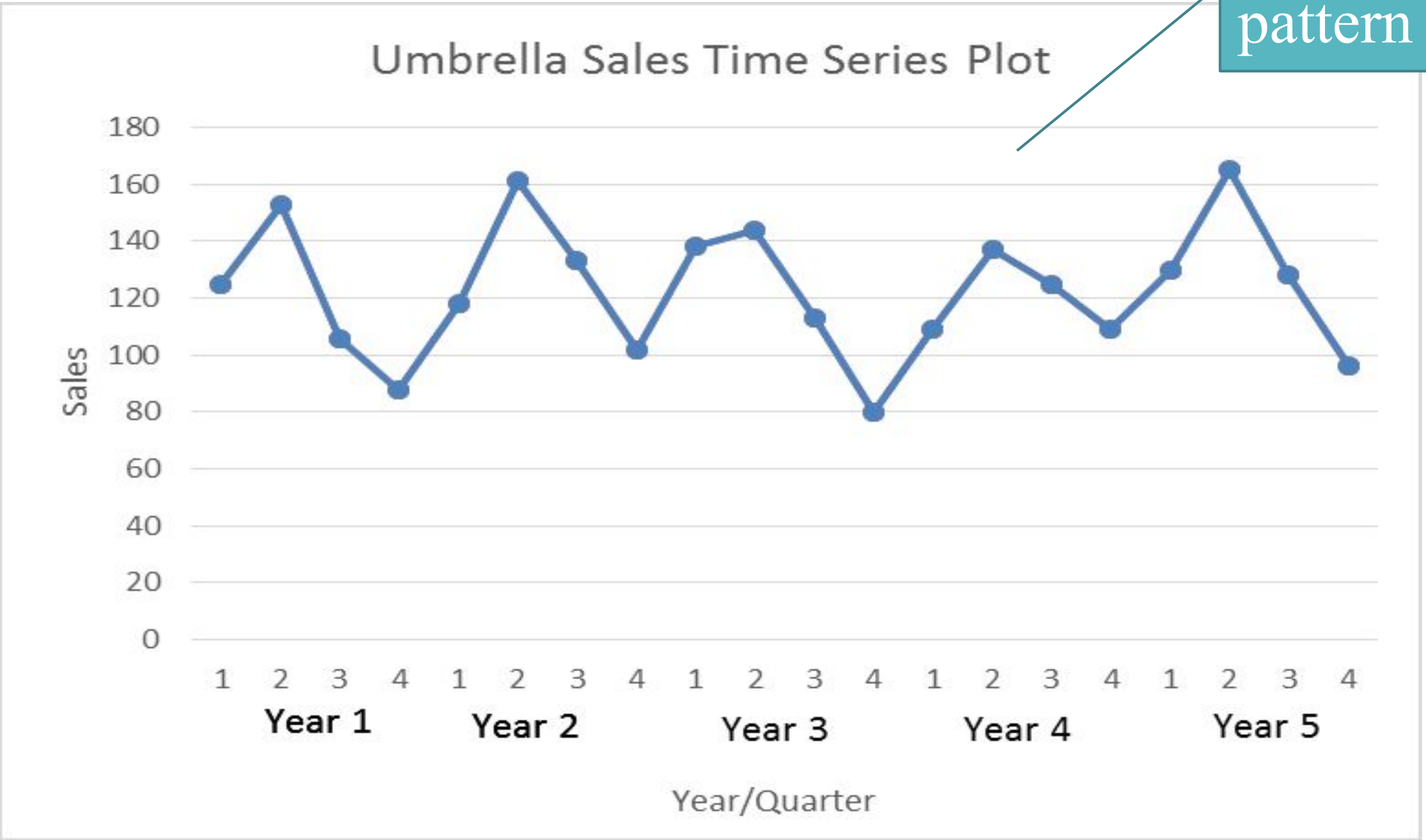
Umbrella Sales Example
우산세일 계절성 예측 예제

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 보십시오.

Time Series Plot (시계열 라인차트)

A quarterly seasonal pattern is present!



Simple Forecasting 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.

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

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

Forecast 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

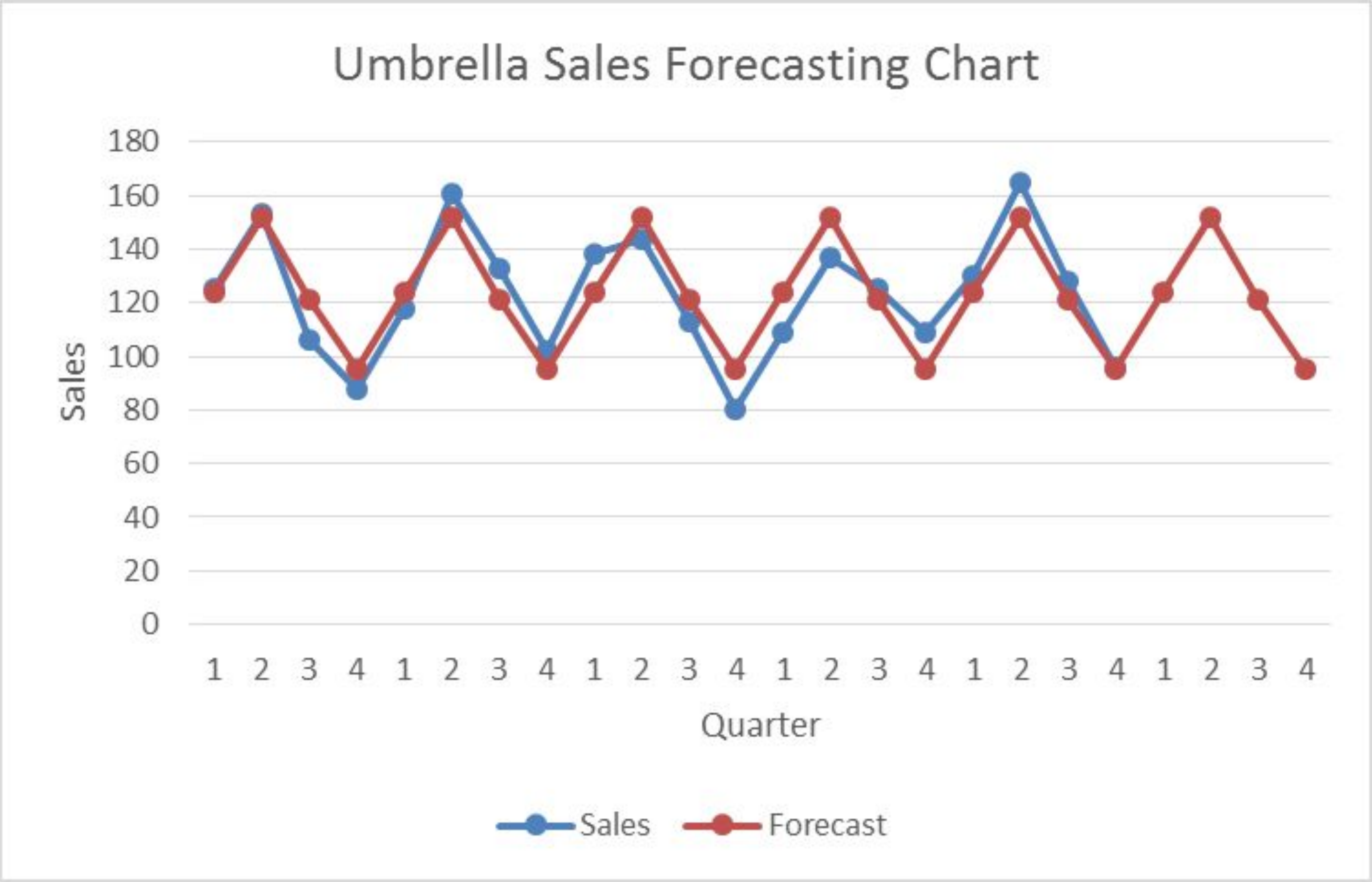
$= 95.0 + 29.0 * 1 + 57.0 * 0 + 26.0 * 0$

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

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

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

Forecasting 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

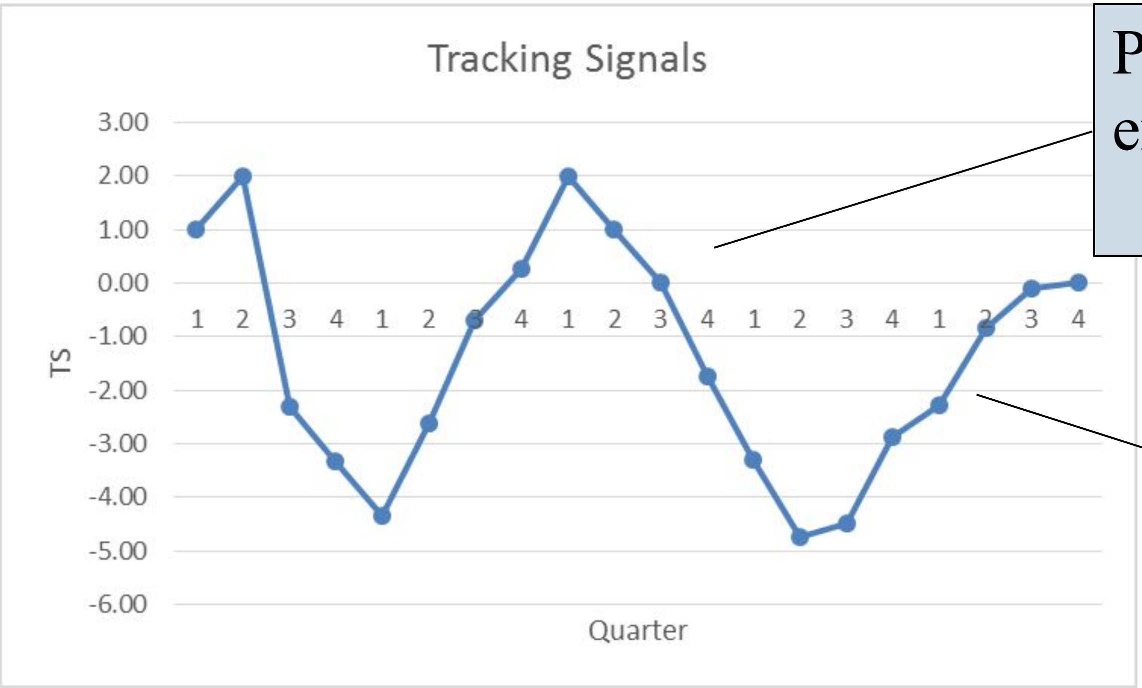
- Error = Sales – Forecast
- ABS Error = ABS(Error)
- % Error = ABS Error / Sales
- Sq. Error = Error^2
- ME = mean(Errors)
- MAE = mean(ABS Errors)
- MAPE = mean(% Errors)
- MSE = mean(Sq. Errors)

Tracking Signals (추적신호)

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

- 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.

Television Set Sales Example TV세일 계절성 추세예측 예제

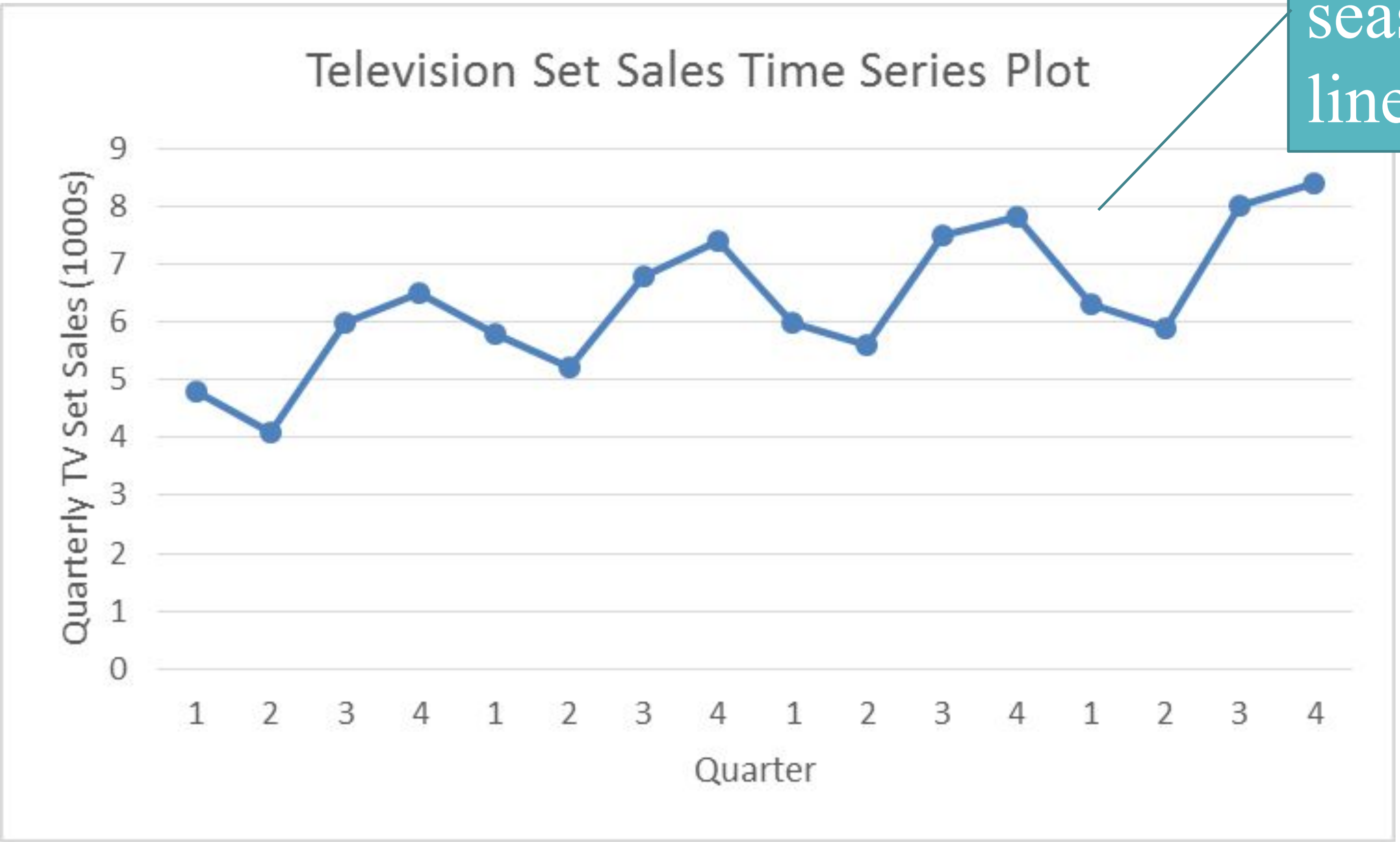
Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8
	4	8.4

The table shows the television set sales for a particular manufacturer over the past four years. Forecast the quarterly television set sales for 5^{th} year.

지난 4년간 특정 제조업자의 티비세트 세일을 보여주고 있습니다. 내년 (Year 5) 이 됐을때 쿼터별 세일을 예측해 보십시오.

Time Series Plot (시계열차트)

The plot contains both a seasonal effect and a linear trend.



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
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

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9880659
R Square	0.9762743
Adjusted R Square	0.9676468
Standard Error	0.2166638
Observations	16

ANOVA

	df	*SS*	*MS*	
Regression	4	21.248	5.312	113
Residual	11	0.516375	0.0469432	
Total	15	21.764375		

	Coefficients	*Standard Error*	*t Stat*	*P-*
Intercept	6.06875	0.162497815	37.346656	6.1
Qrt1	-1.363125	0.157454336	-8.657272	3.
Qrt2	-2.03375	0.155107642	-13.11186	4.6
Qrt3	-0.304375	0.153682427	-1.980545	0.
Period	0.145625	0.012111872	12.023327	1.

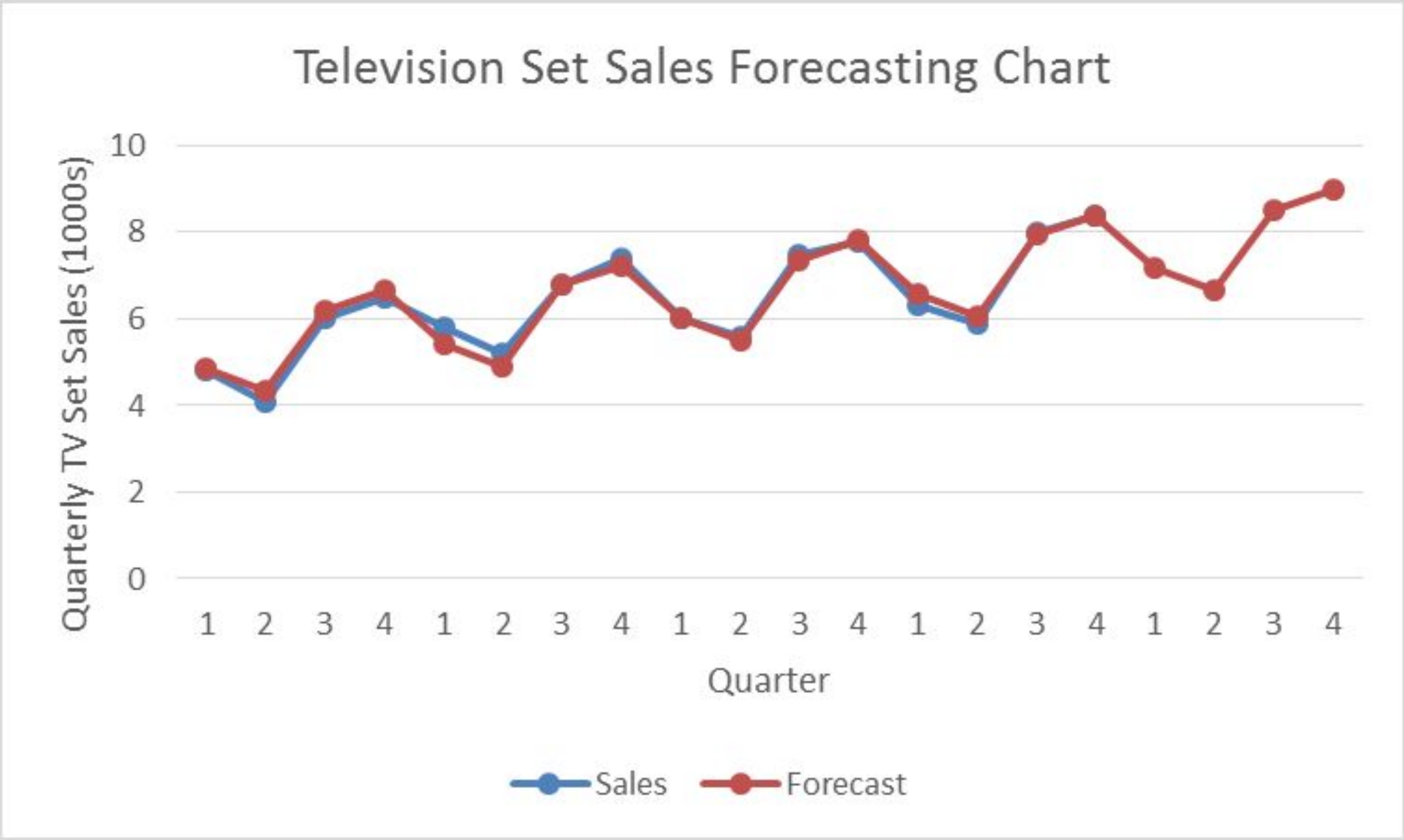
$= 6.07 - 1.36 * 1 - 2.03 * 0 - .304 * 0 + .146 * 17$

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

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

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

Forecasting 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

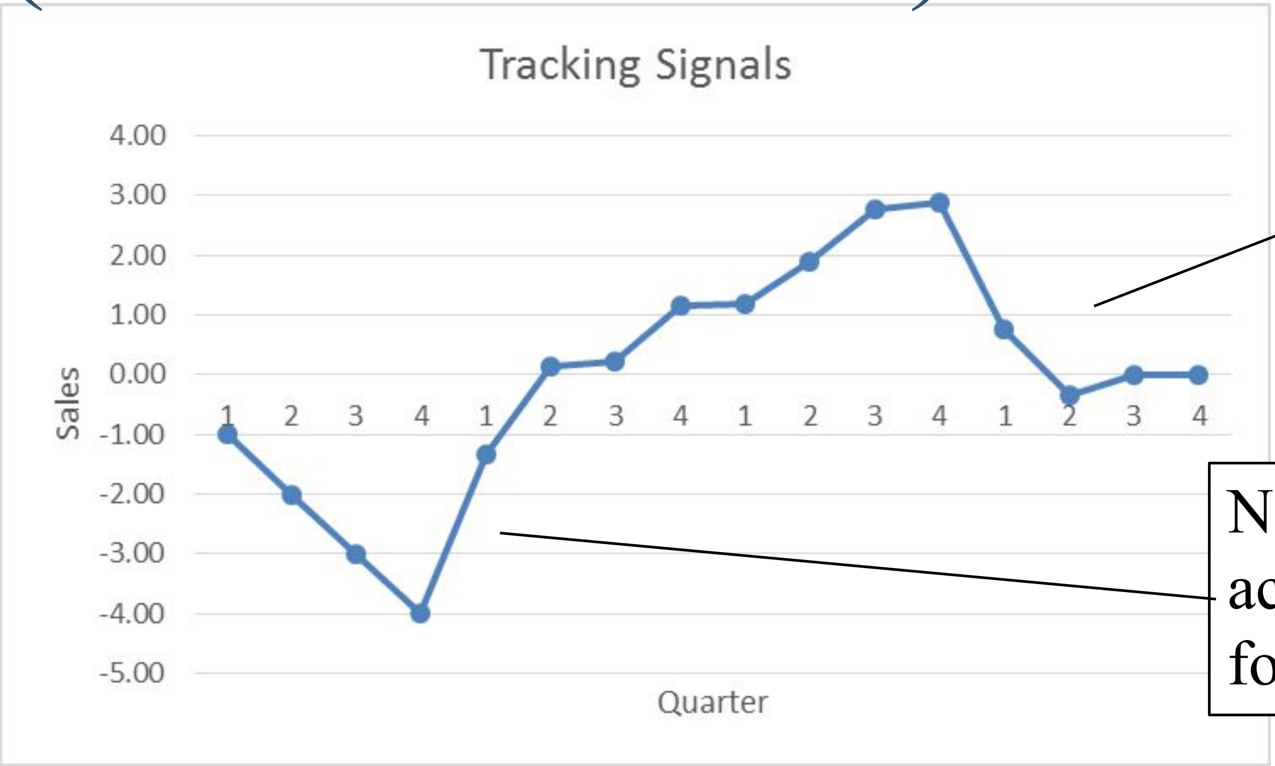
- Error = Sales – Forecast
- ABS Error = ABS(Error)
- % Error = ABS Error / Sales
- Sq. Error = Error^2
- ME = mean(Errors)
- MAE = mean(ABS Errors)
- MAPE = mean(% Errors)
- MSE = mean(Sq. Errors)

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

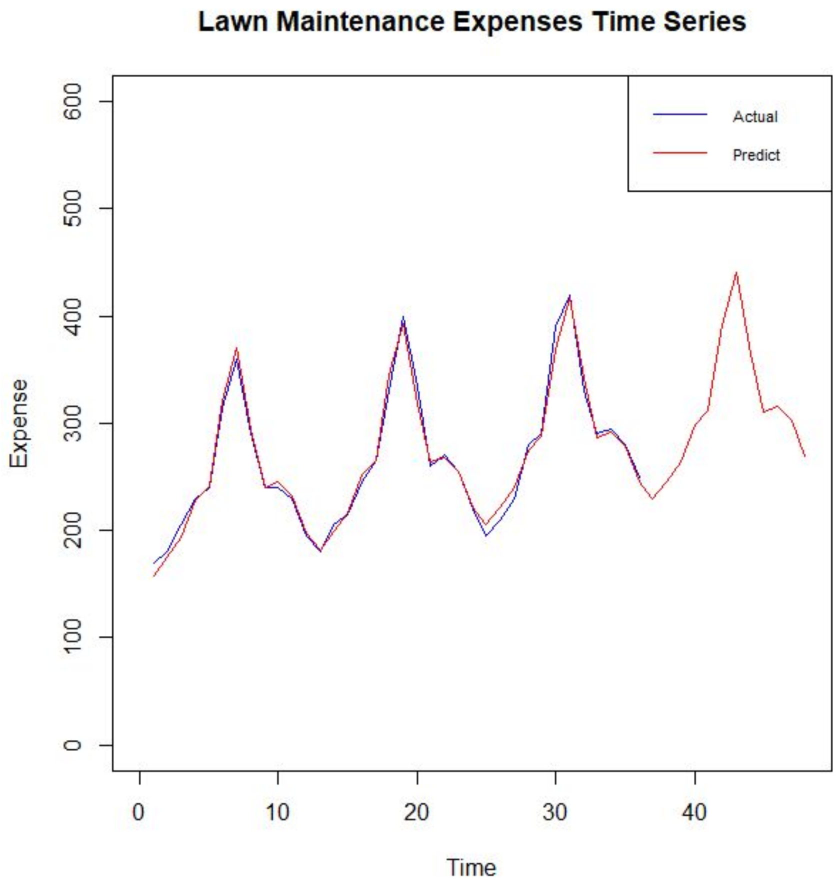
As long as TS is between -4 and 4, assume the model is working correctly.

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.

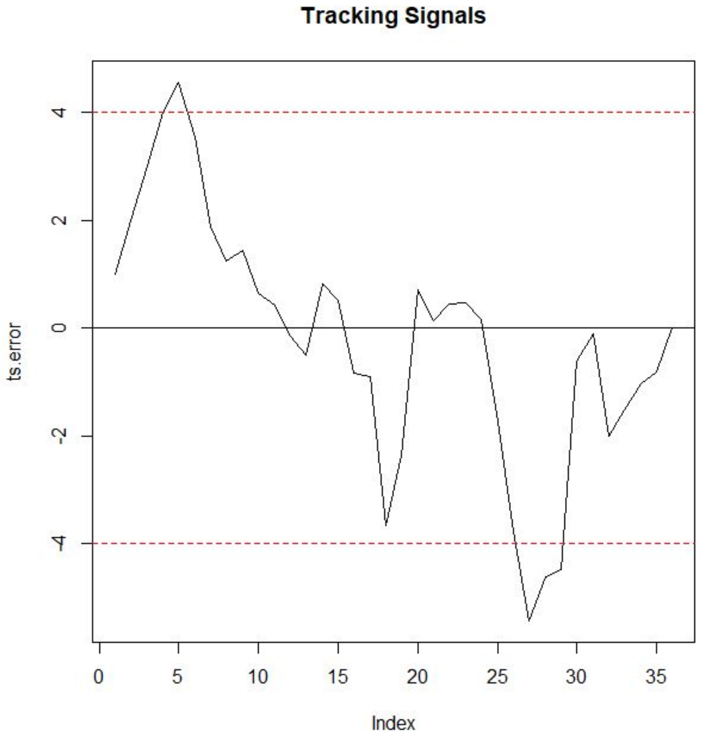
Forecast Sales for 4th Year (예측값)




```
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 #8

- Umbrella Sales Example
  - Plot the time series
  - Conduct a seasonal forecast and draw the time series plot
  - Forecast year 11 and year 12
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

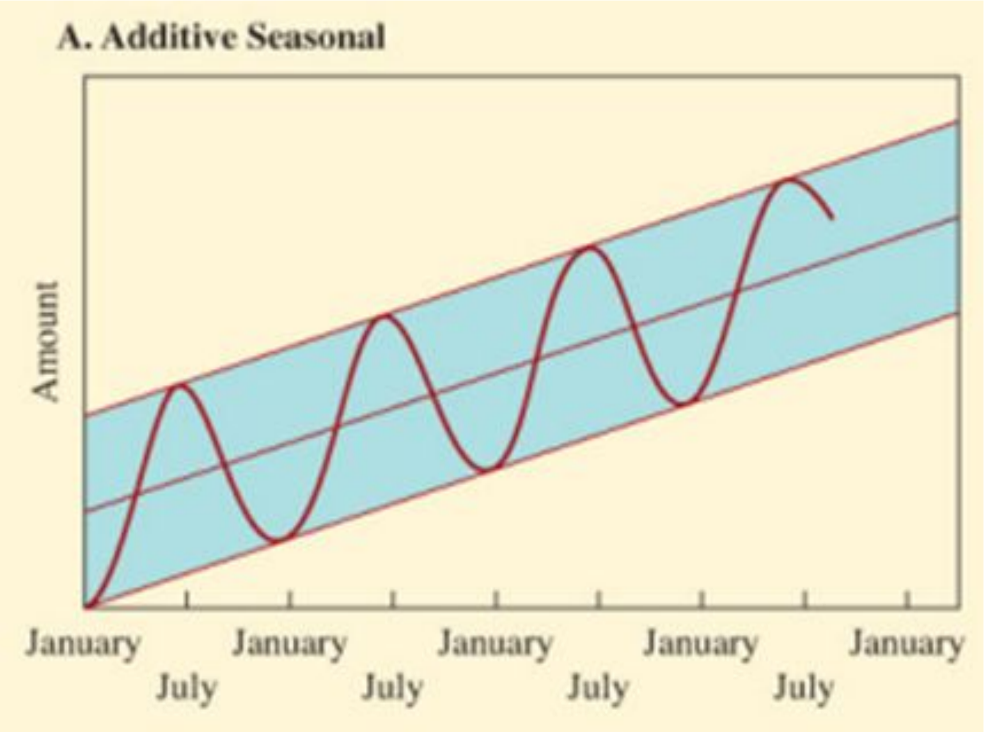
Exercise #8

- TV Set Sales Example
  - Plot the time series
  - Conduct a seasonal forecast and draw the time series plot
  - Forecast year 11 and year 12
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

Exercise #8

- Lawn-Maintenance Expense Example
  - Plot the time series
  - Conduct a seasonal forecast and draw the time series plot
  - Forecast year 11 and year 12
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

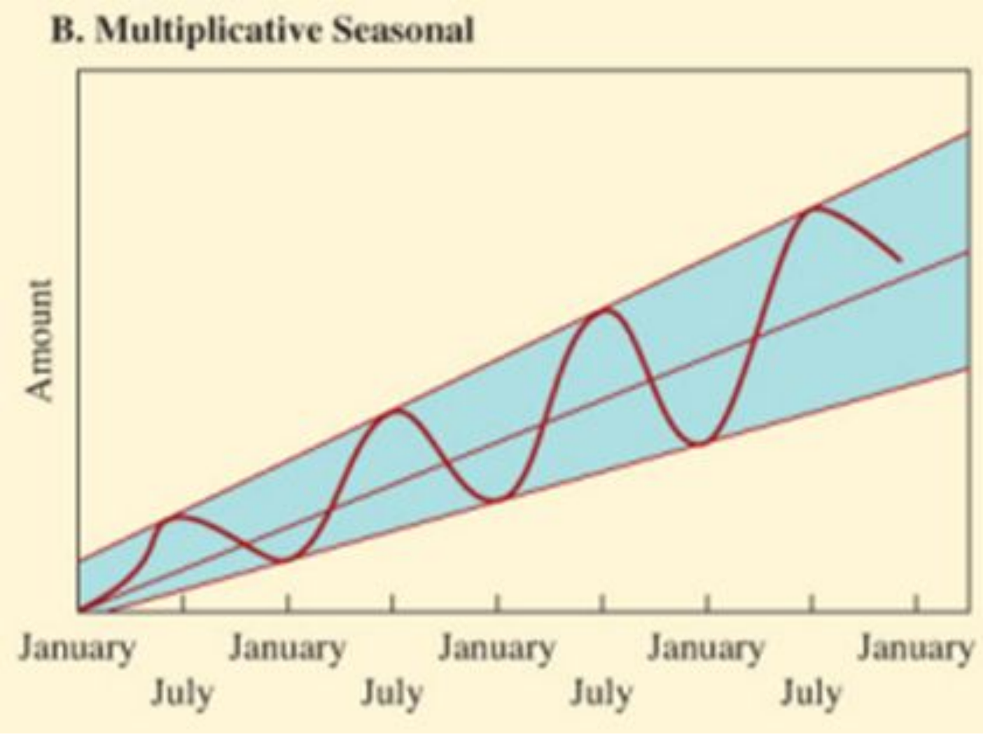
Additive Seasonal Forecasting (가법계절예측)



- Additive seasonal variation assumes that seasonal amount is a constant no matter what the trend or average amount is.

- Additive seasonal forecasting = trend + seasonal amounts

Multiplicative Seasonal Forecasting (승법계절예측)



- In Multiplicative seasonal variation, the trend is multiplied by the seasonal factors.

- Multiplicative seasonal forecasting = trend x seasonal factor

Time Series Components (시계열 요소들)

- A seasonal time series consists of a trend component, a seasonal component and an irregular component.

Trend

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

Cycles

- Patterns in the data that occur every several years

Seasonality

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

Random Variations

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

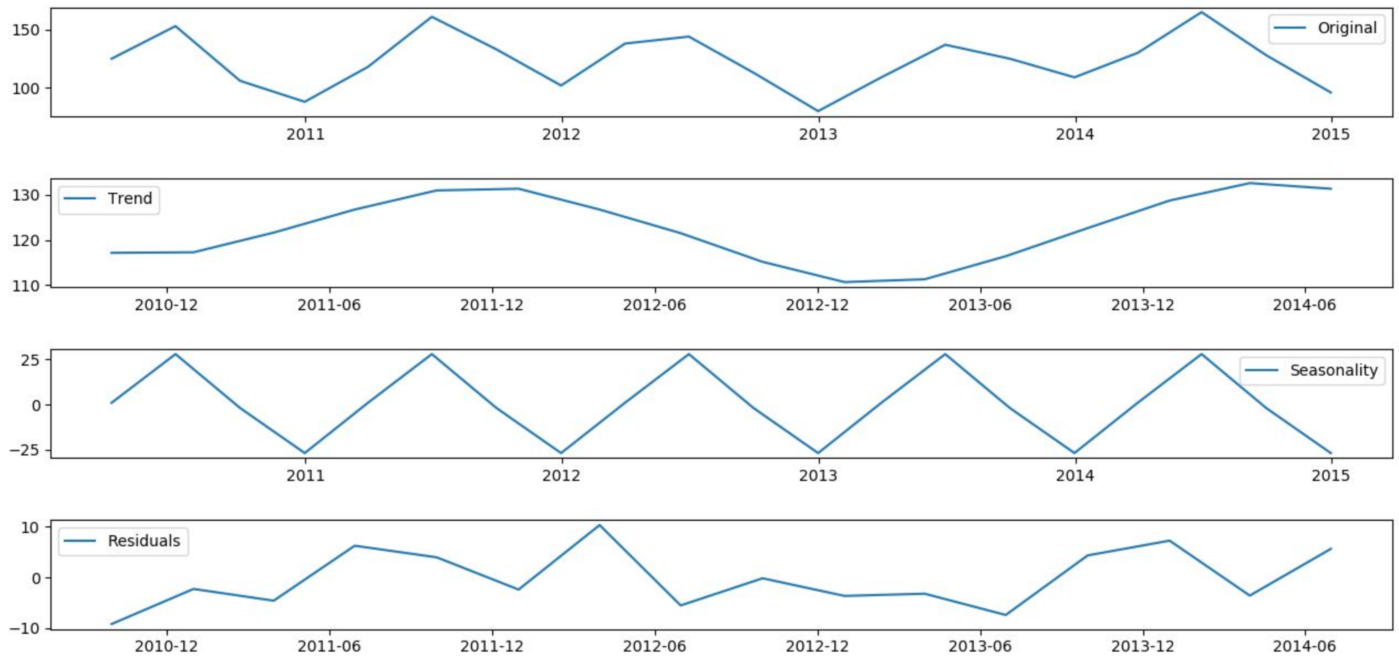
Decomposing Seasonal Data (계절요소의 분해)

- Separating the time series into seasonal, trend, and random components.

```
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()
```



Season Adjusted Plot

```
ts_adj = ts - seasonal
plt.plot(ts_adj)
plt.ylim(0,200)
x = np.arange(1,21)
y = ts_adj

import statsmodels.api as sm
x = sm.add_constant(x)
model = sm.OLS(y,x).fit()
model.summary()
```

Composition

```
ts_adj_pred = model.predict(x)
ts_pred = ts_adj_pred + seasonal
plt.plot(dt,ts_adj_pred)
plt.plot(dt,ts_pred)
plt.ylim(0,200)
```

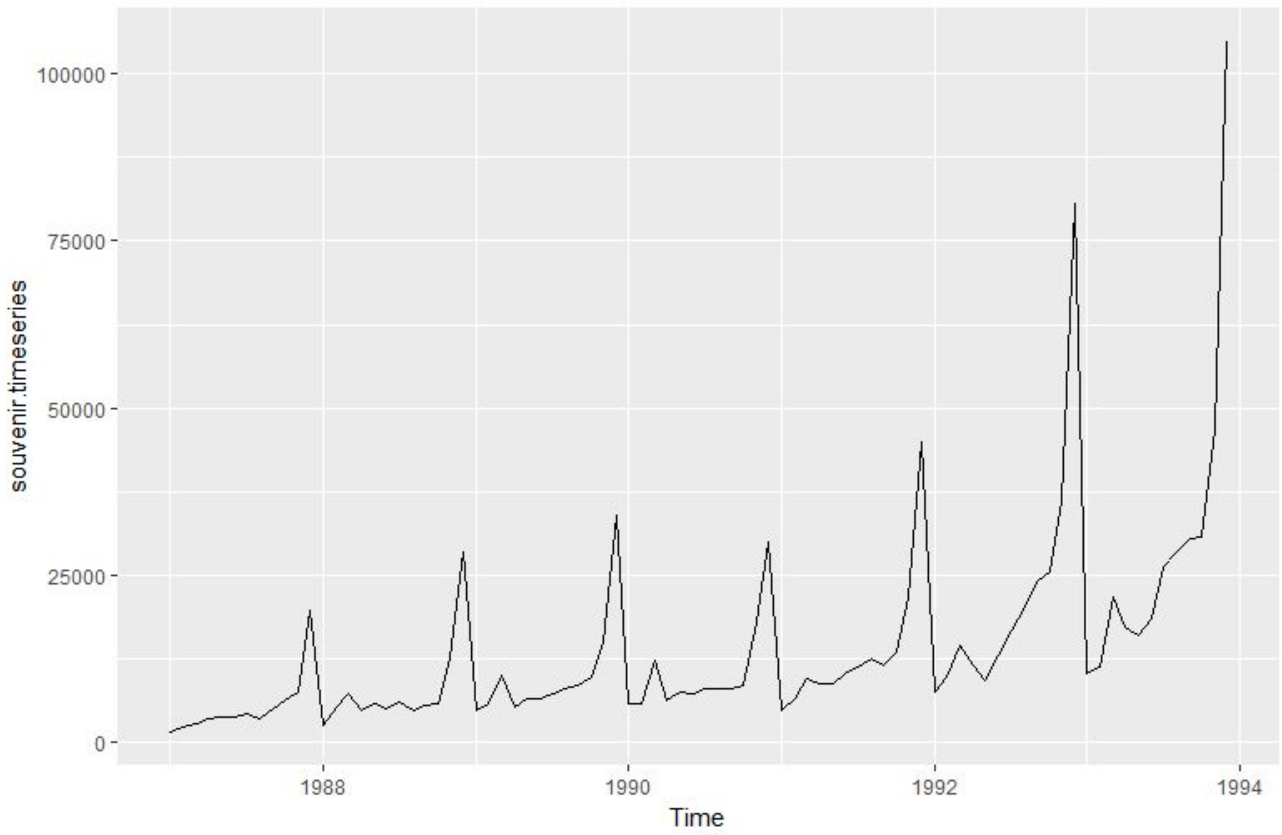
Souvenir Shop Sales Example
기념품가게 승법계절예측

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 예측하십시오.

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.

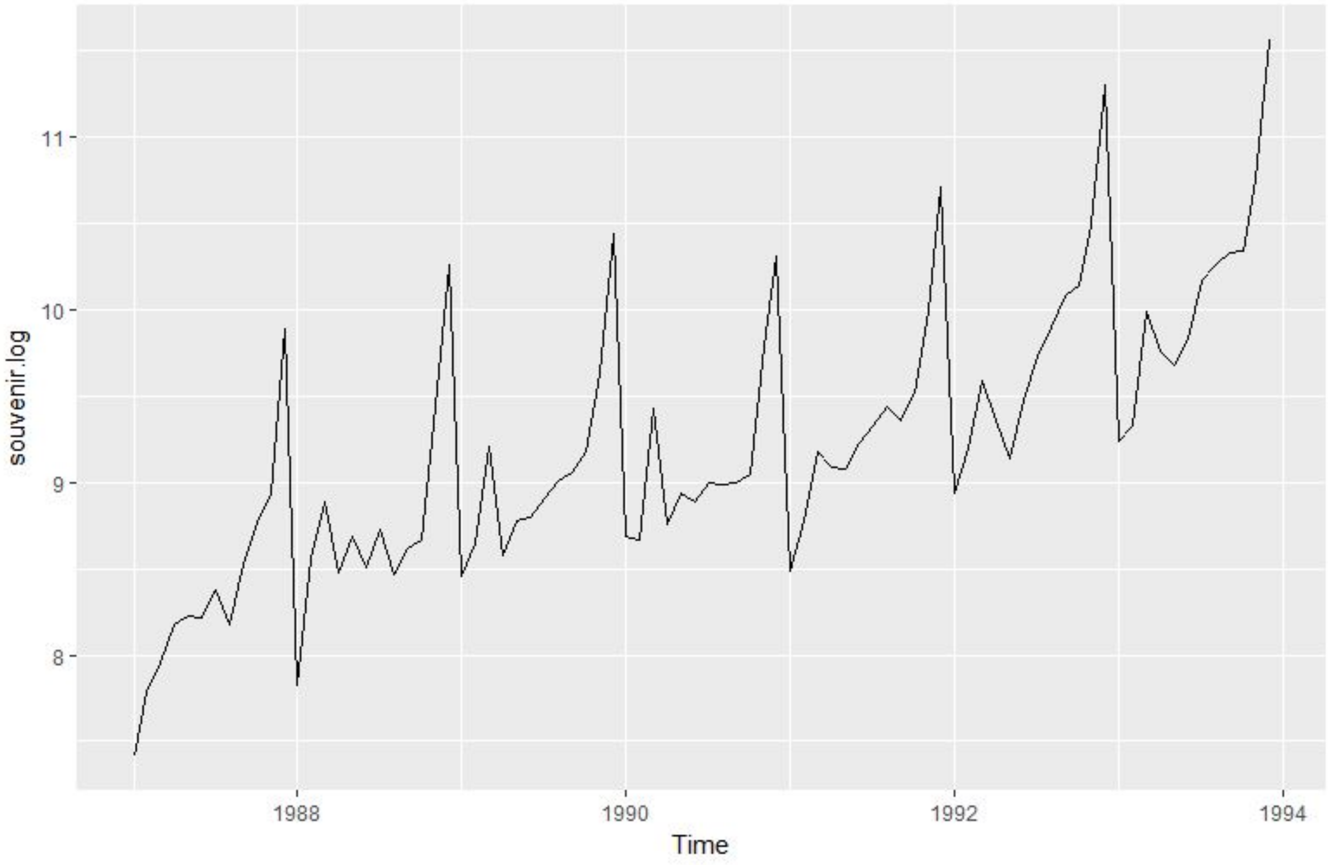


Log Transformation (로그변환)

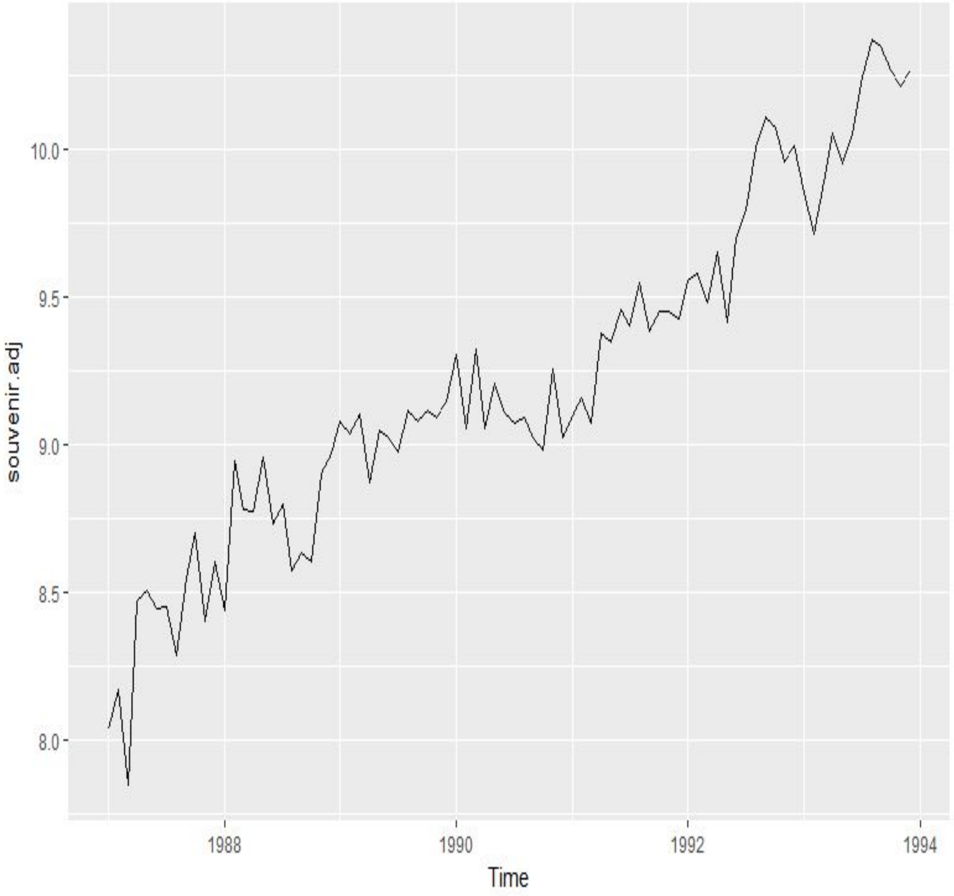
```
souvenior_log =
np.log(souvenior)
```

```
plt.plot(souvenior_log)
```

- The log-transformed time series can be described using an additive model.

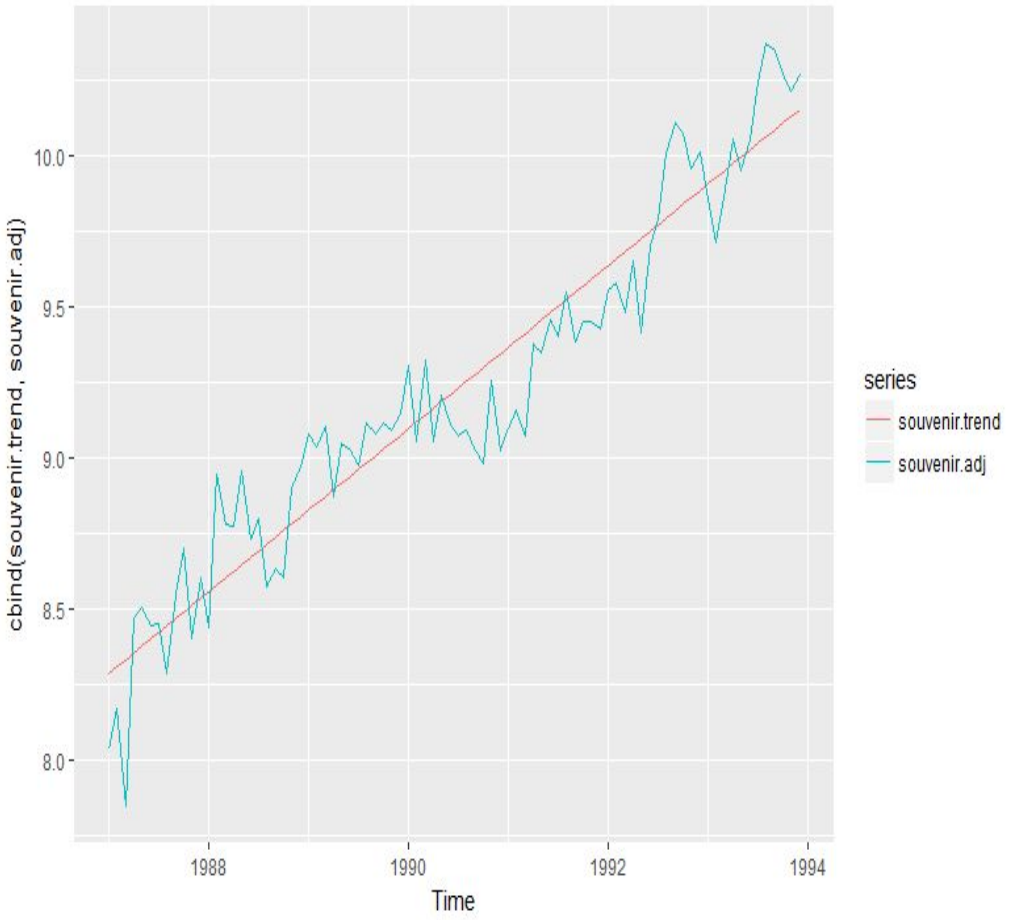


Seasonal Decomposition (계절분해)



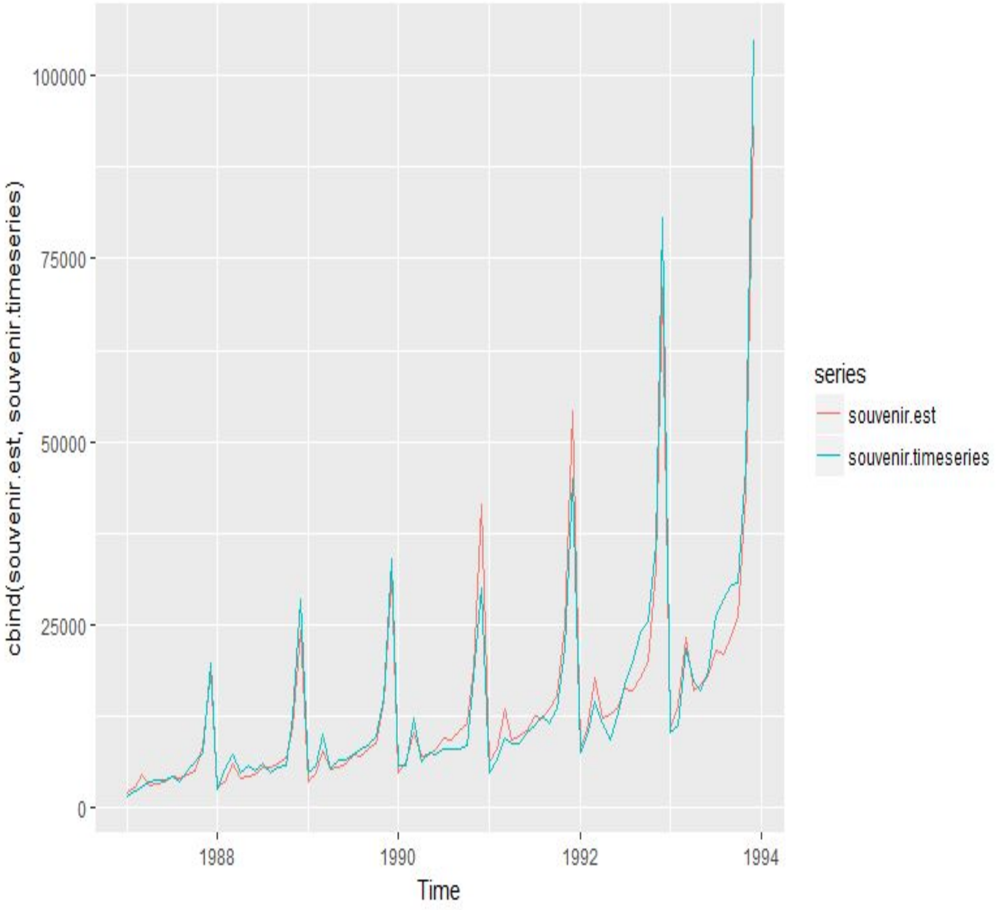

```
decomposition =
seasonal_decompose(souvenio
r_log)

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)
```

Recomposition (복구)




```
souvenior_comp =
souvenior_trend + seasonal

souvenior_pred =
np.exp(souvenior_comp)

plt.plot(souvenior)

plt.plot(souvenior_pred)
```

Exercise #8

- Souvenir Example
  - Import the souvenir data and transform using log. Decompose the seasonality and conduct the regression analysis. Recompose the data and forecast sales for next 12 months.

Exercise #8

- Umbrella Sales Example
  - Create an umbrella time series (freq='Q') and plot it using plt.plot
  - Forecast year 6 using seasonal decomposition method (subtract seasonal factor □ linear regression □ add seasonal factor)
  - Compare this result to the result from yesterday

Exercise #8

- TV Sets Sales Example
  - Create a tvset time series (freq=‘Q’) and plot it using plt.plot
  - Forecast year 6 using seasonal decomposition method (subtract seasonal factor □ linear regression □ add seasonal factor)
  - Compare this result to the result from yesterday

Exercise #8

- Lawn-Maintenance Expense Example
  - Create a tvset time series (freq='M') and plot it using plt.plot
  - Forecast year 4 using seasonal decomposition method (subtract seasonal factor □ linear regression □ add seasonal factor)
  - Create an error table and calculate ME, MAE, MAPE, and MSE
  - Create a tracking signal table and plot the signals

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
- [x] 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

Exercise #8

- 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 2017. Forecast monthly sales in 2018.
- Furniture Store Sales Example
  - Find furniture store sales from 2010 to 2017 from the US Census and forecast monthly sales in 2018