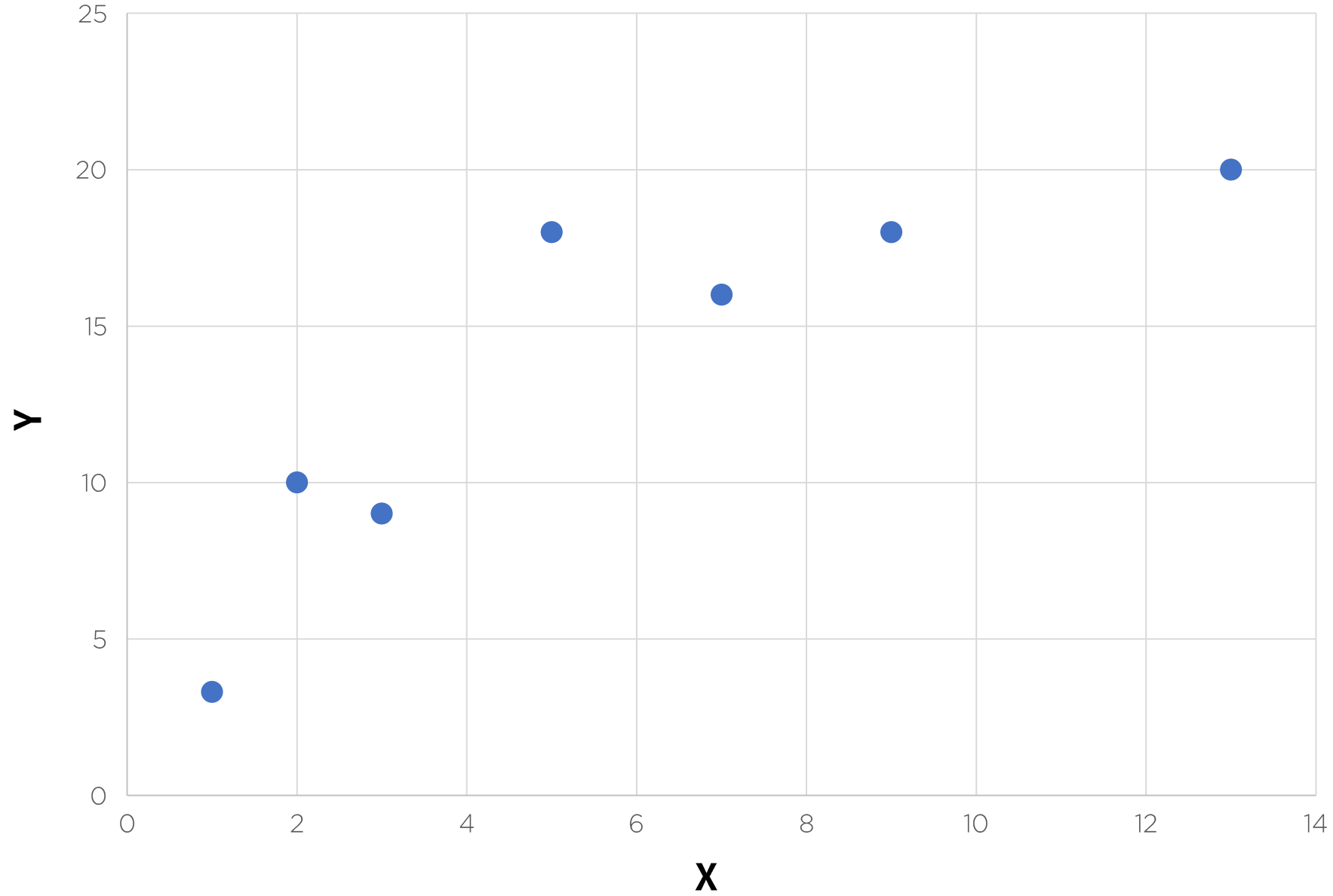
Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

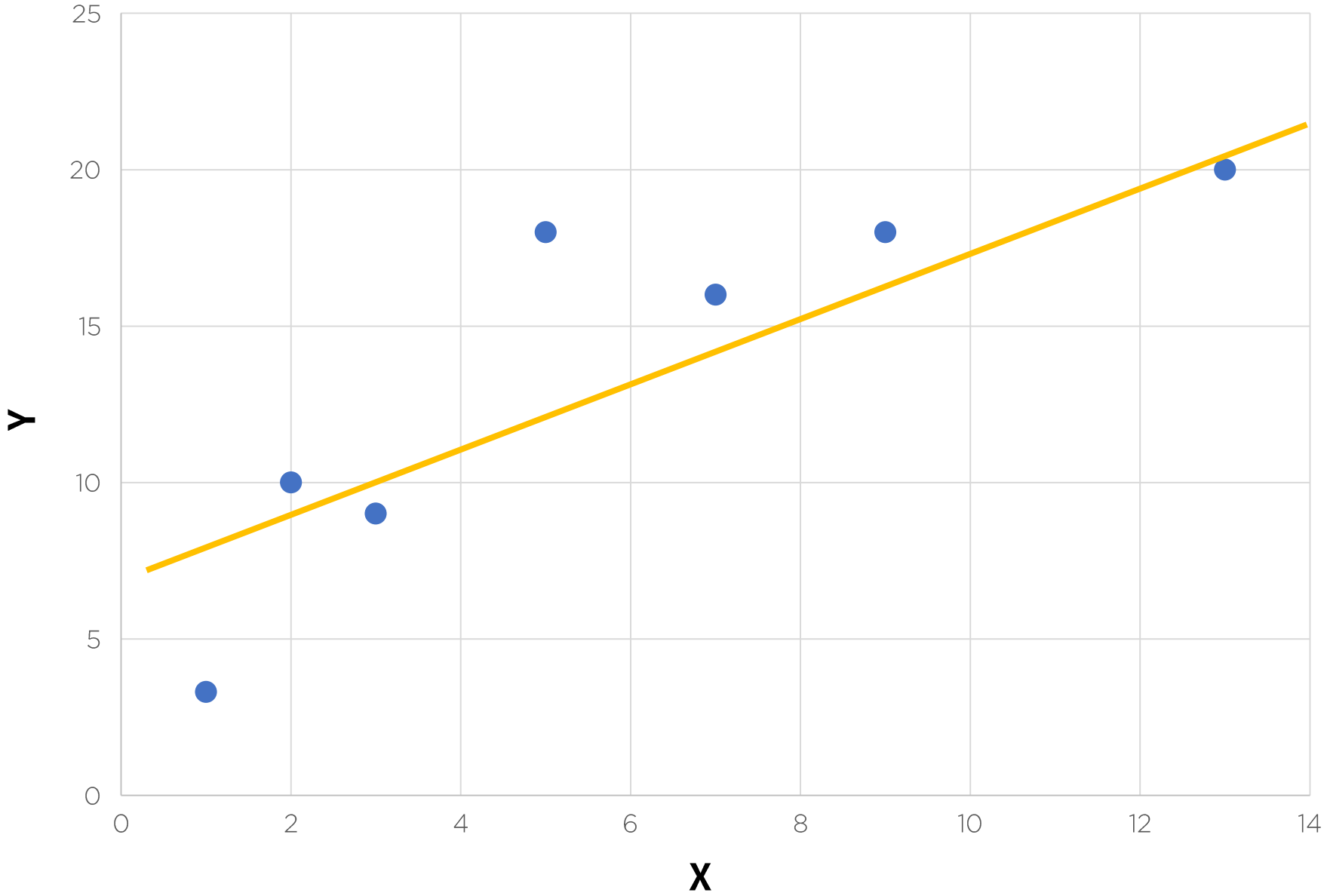
분석의 목적

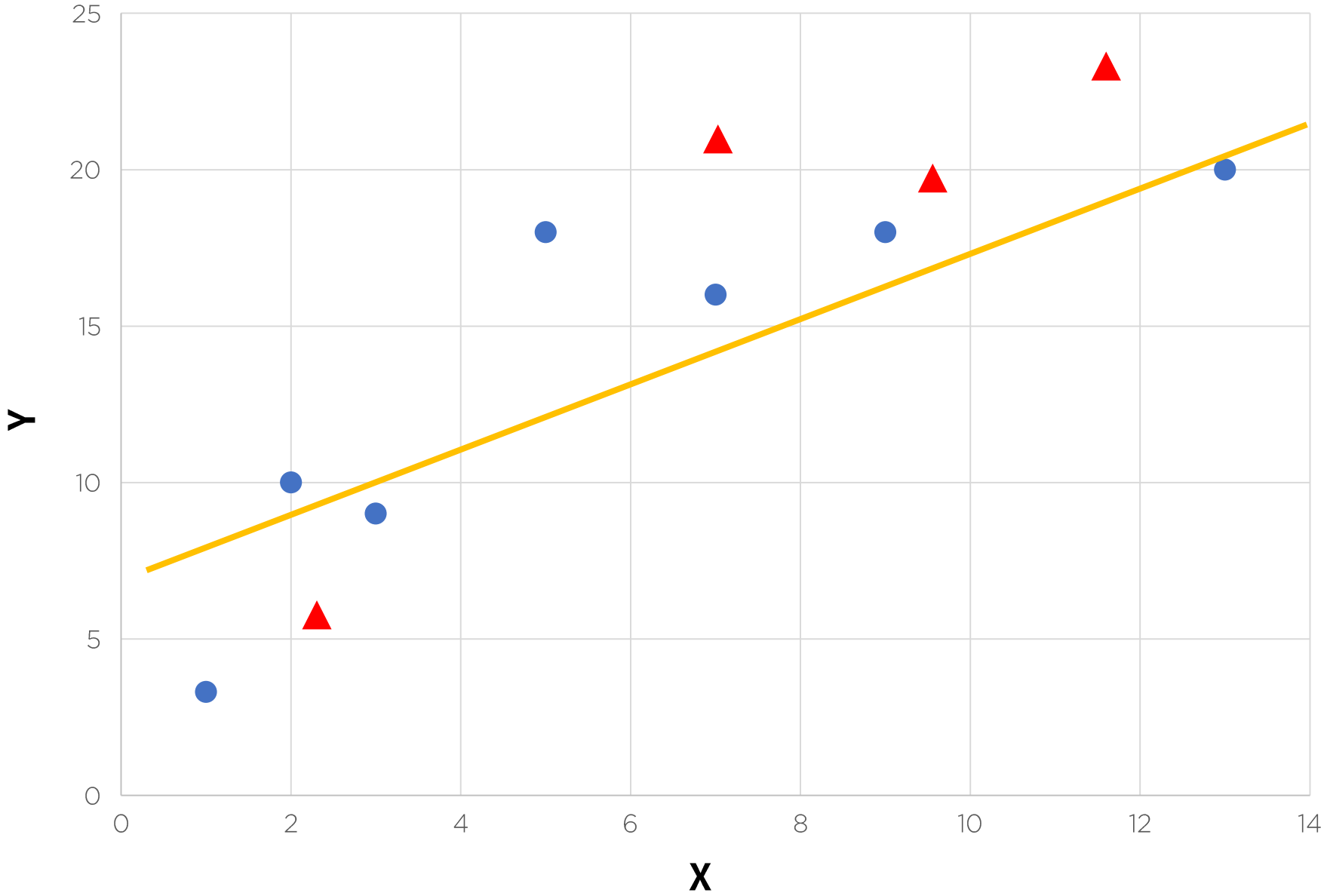
Linear Regression을 사용하여 E-Commerce 고객별 연간 지출액을 예측

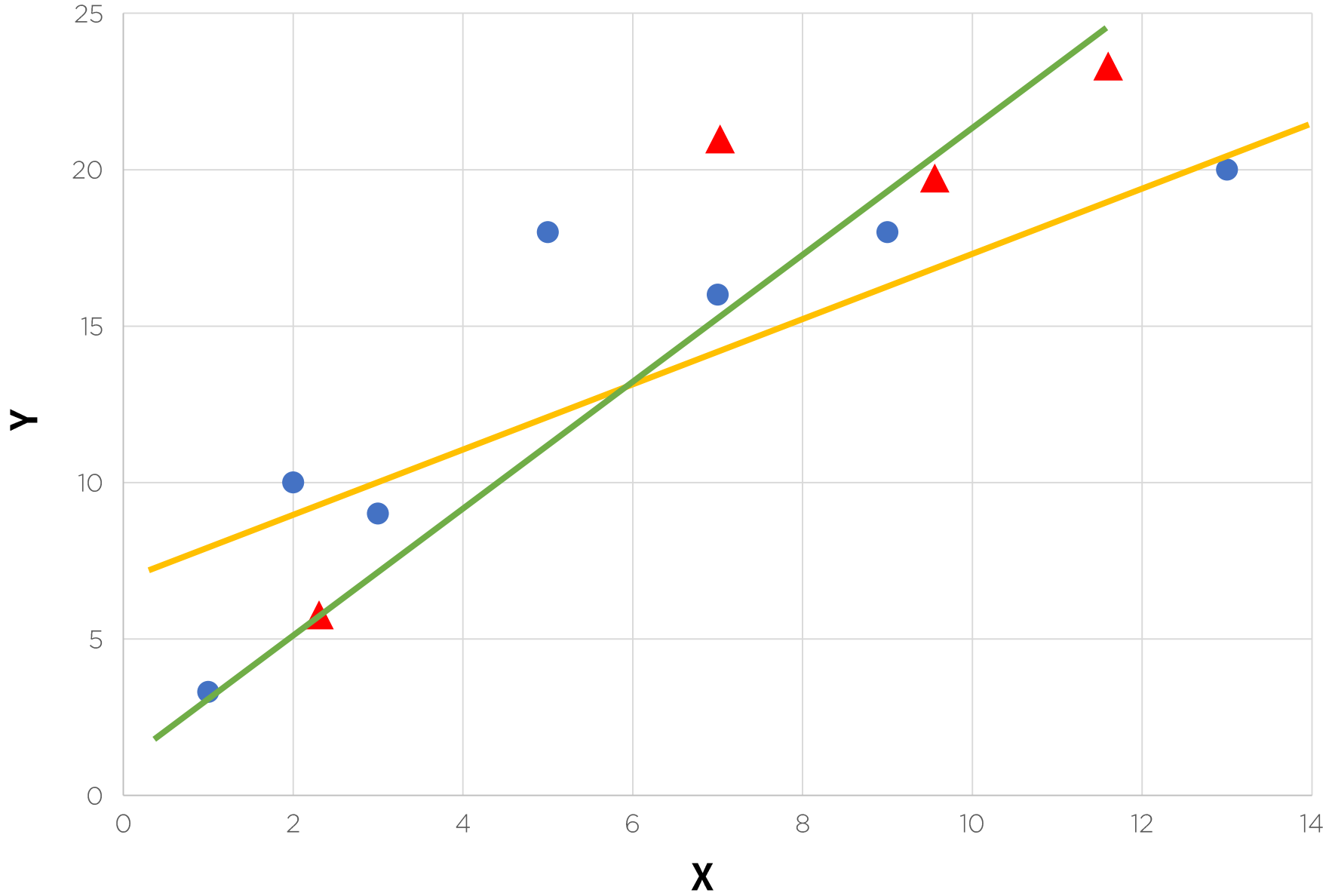
Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

Train Test Split



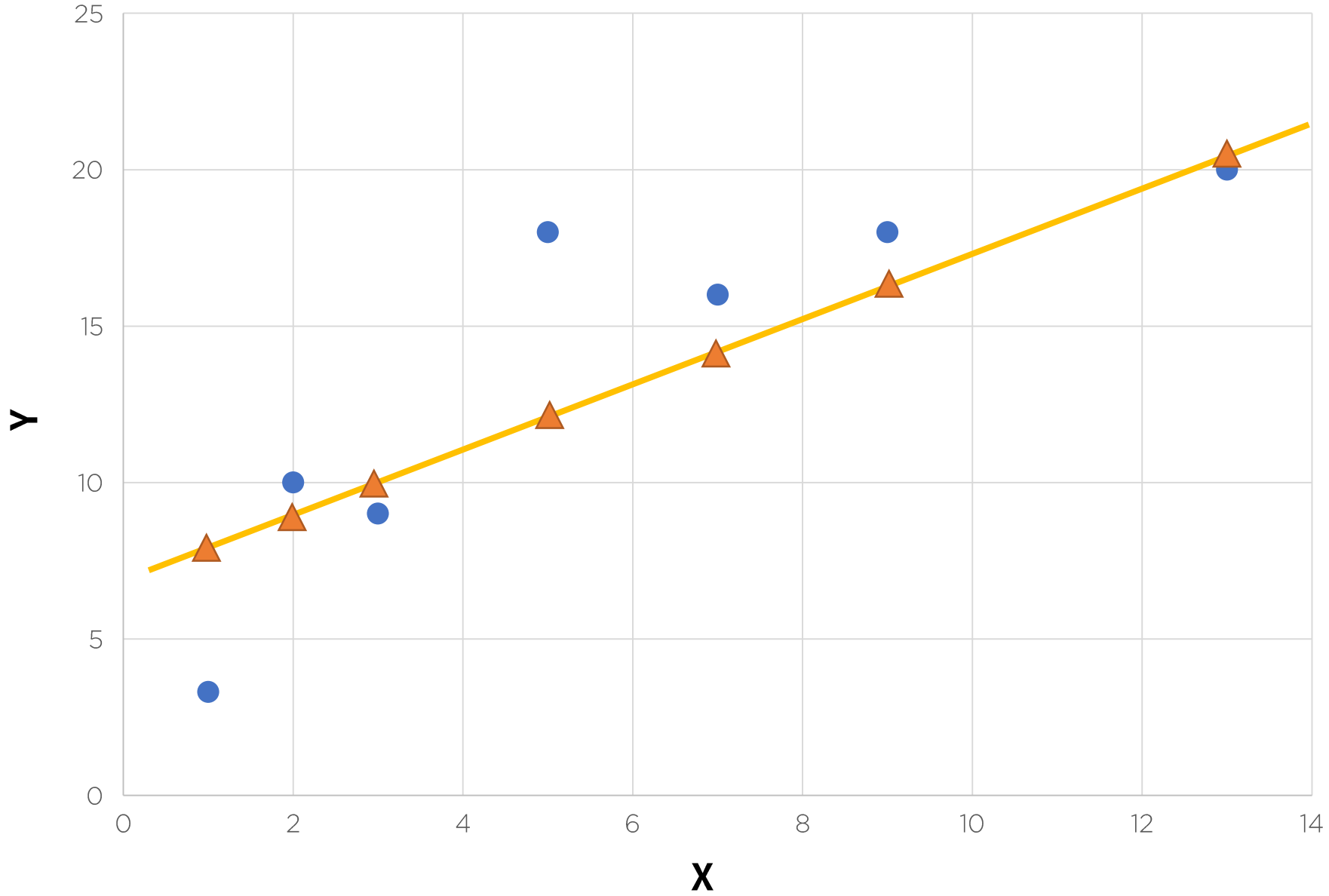


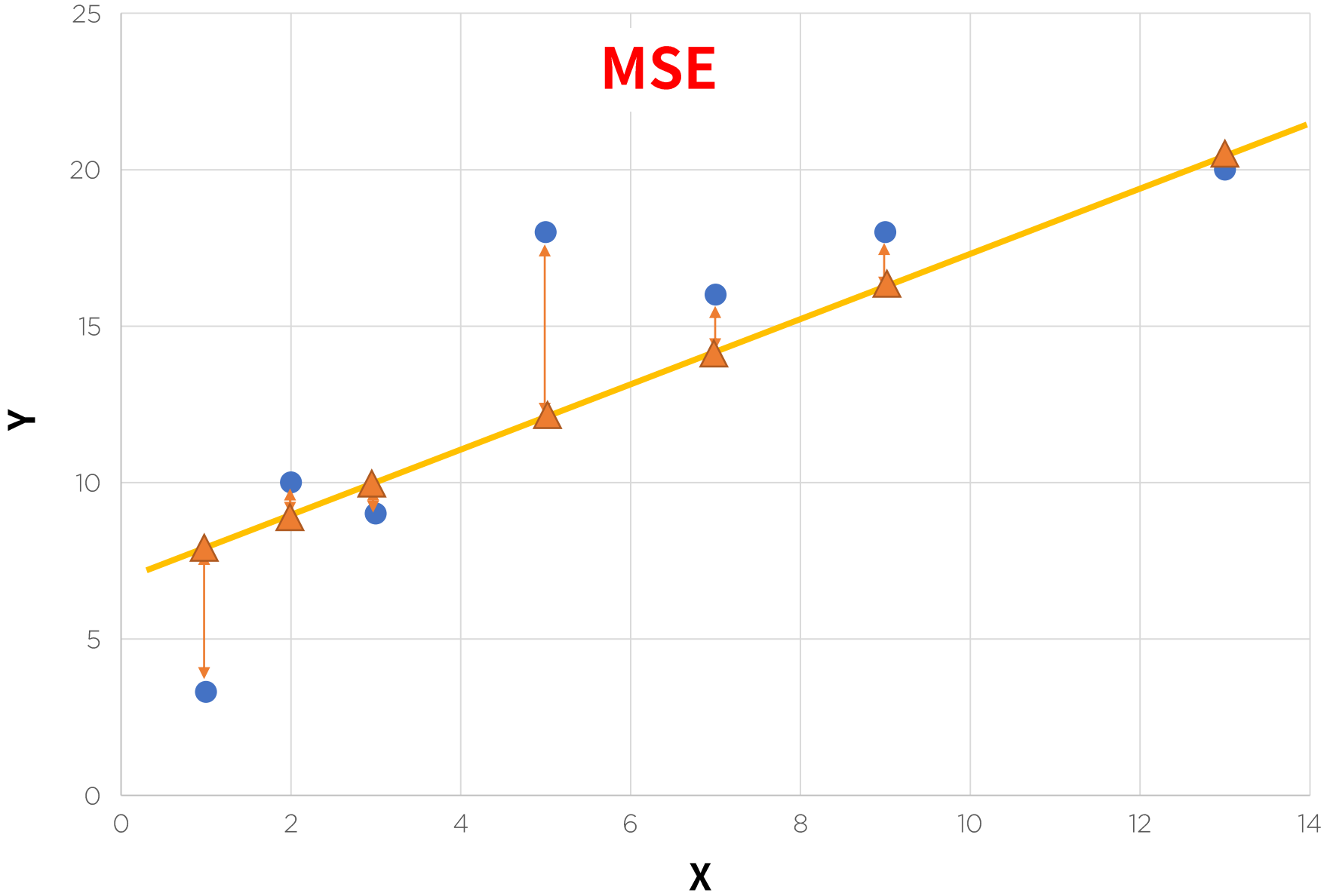




Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

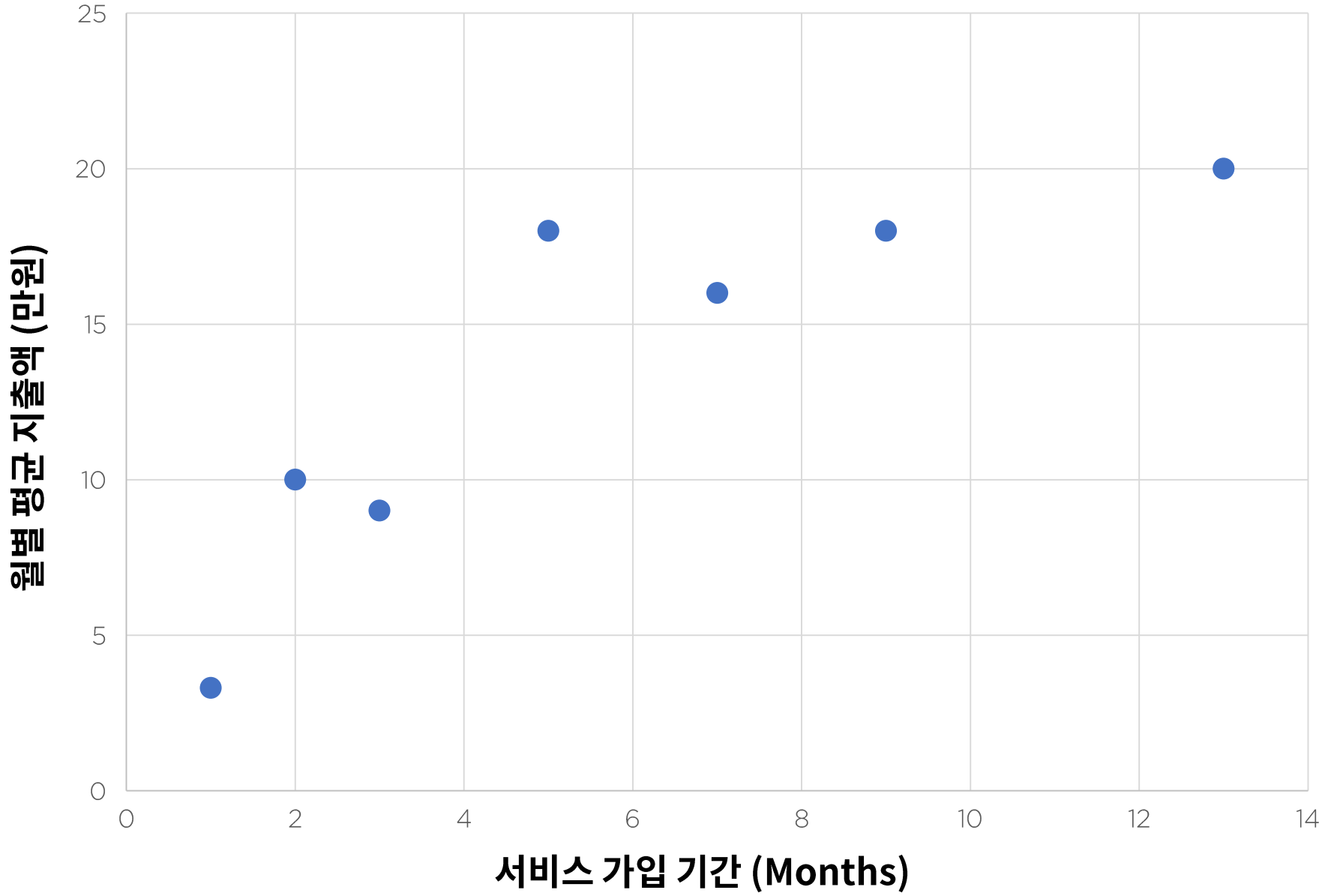
모델을 활용하여 예측하고 평가하기

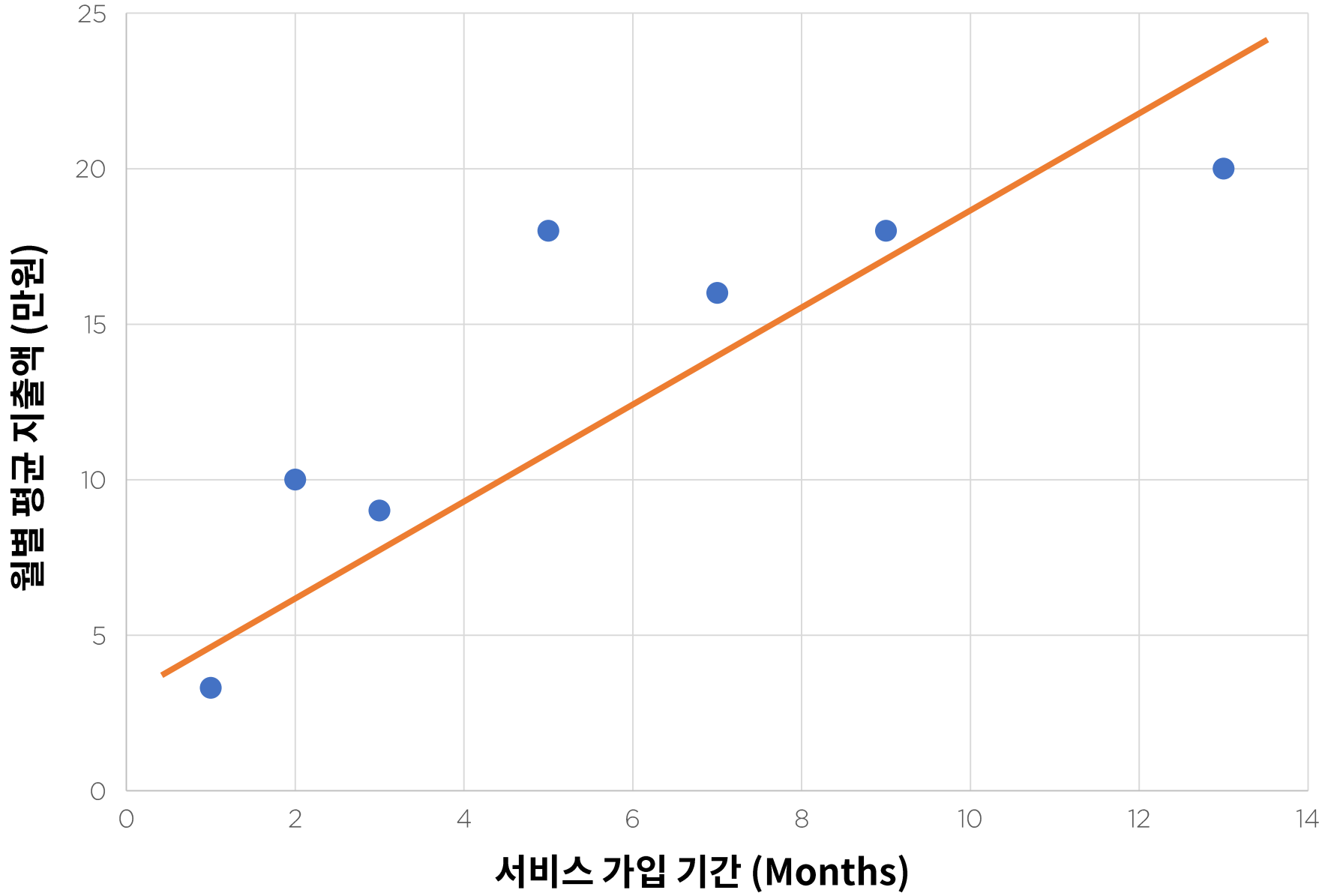


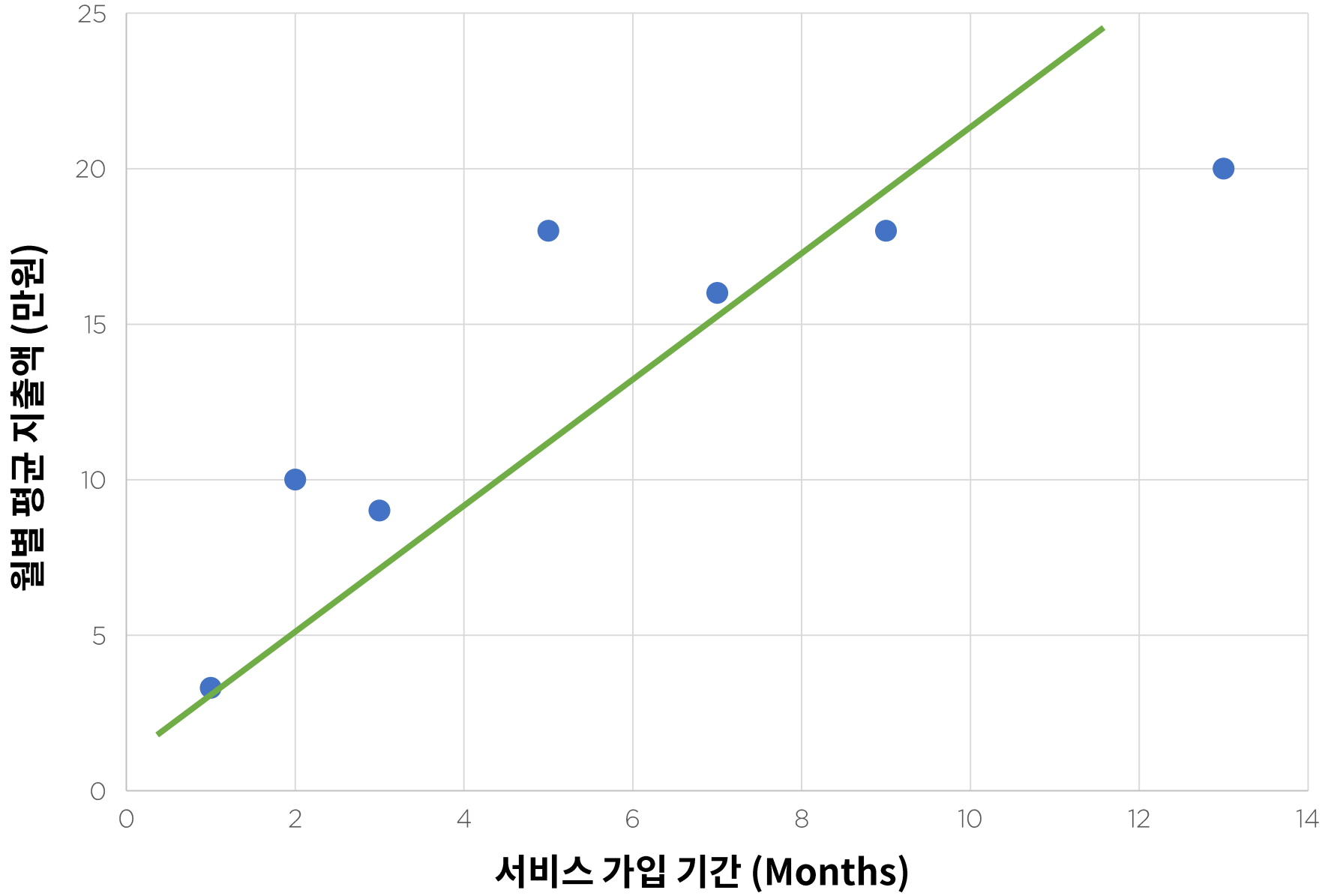


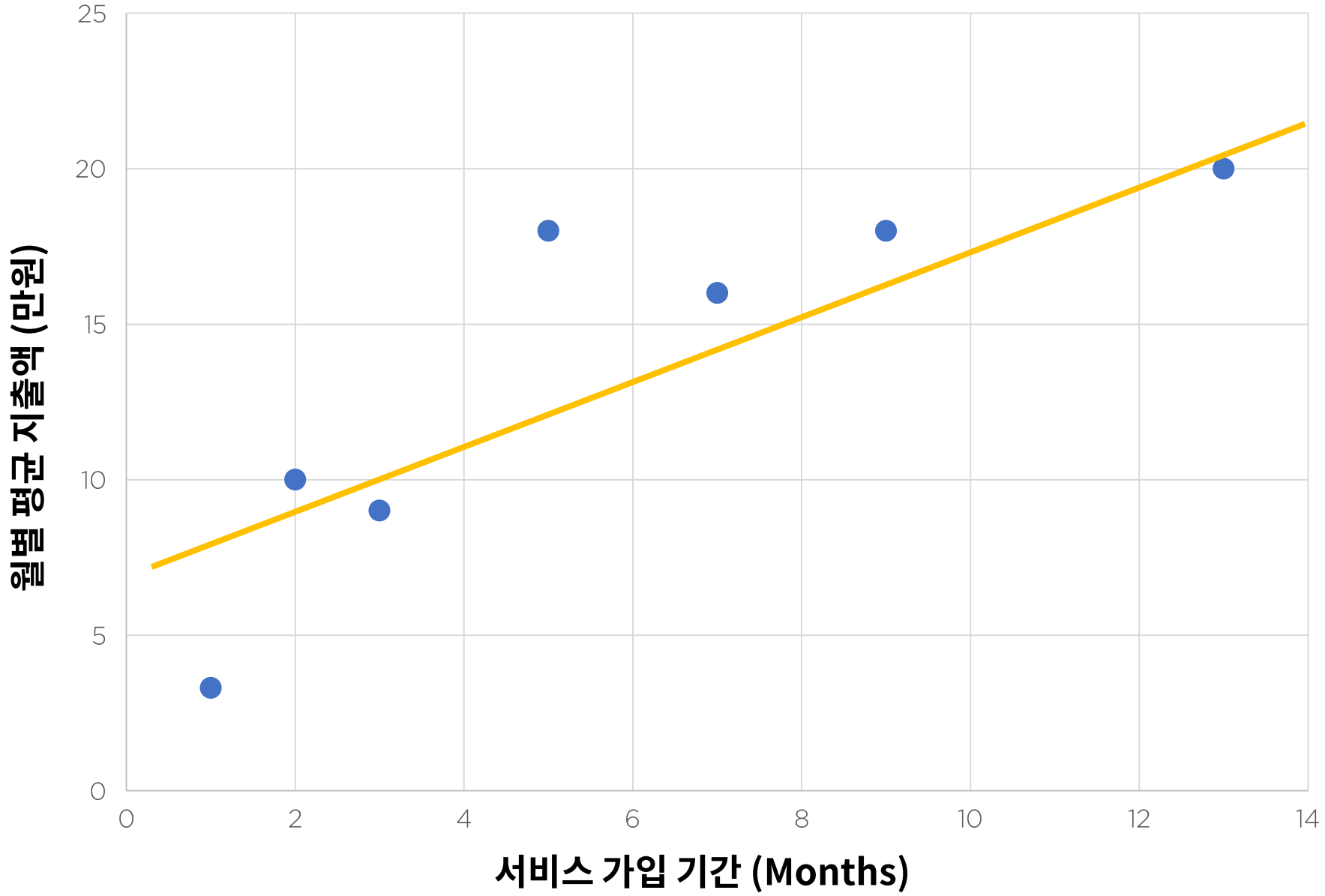
Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

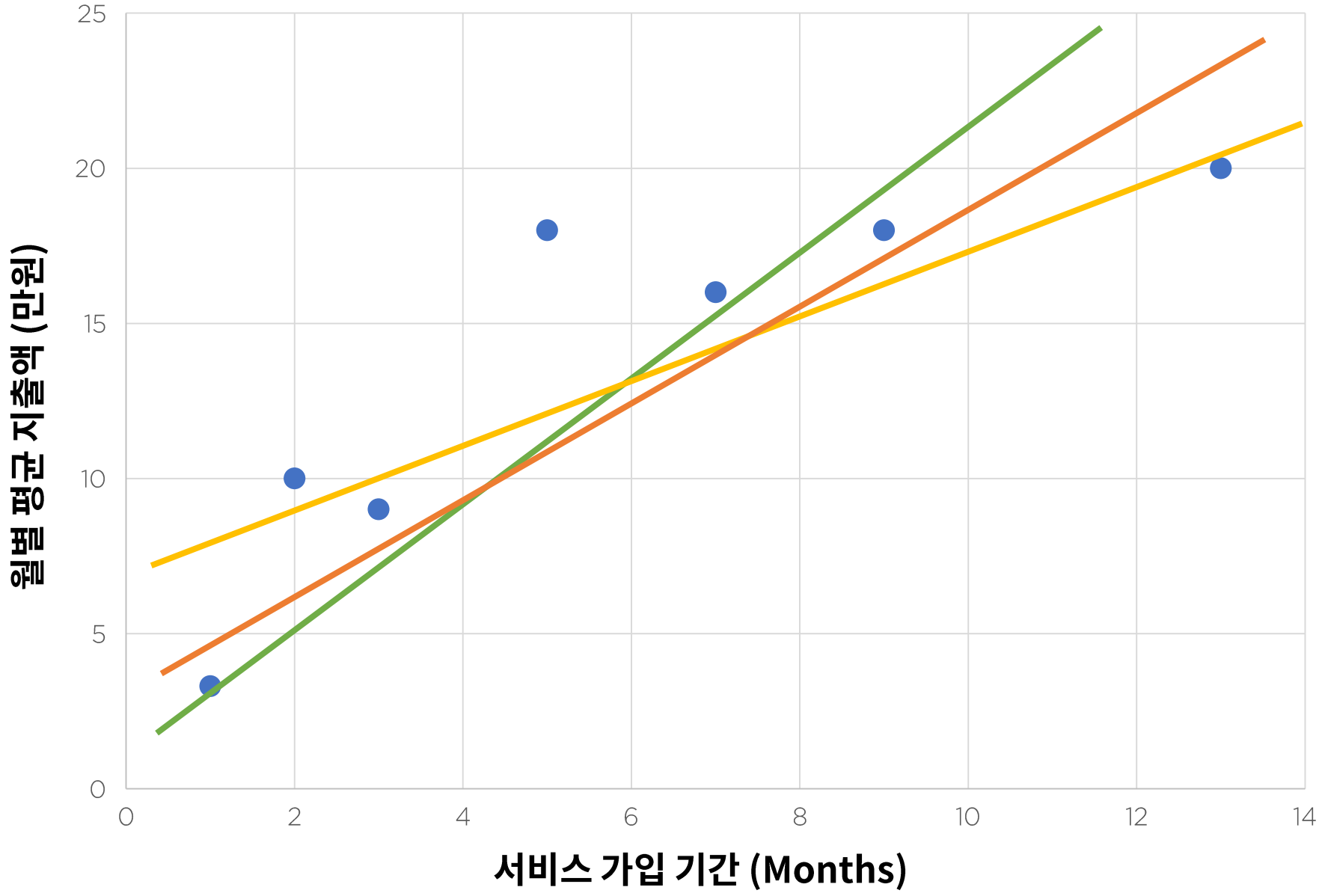
Linear Regression의 원리

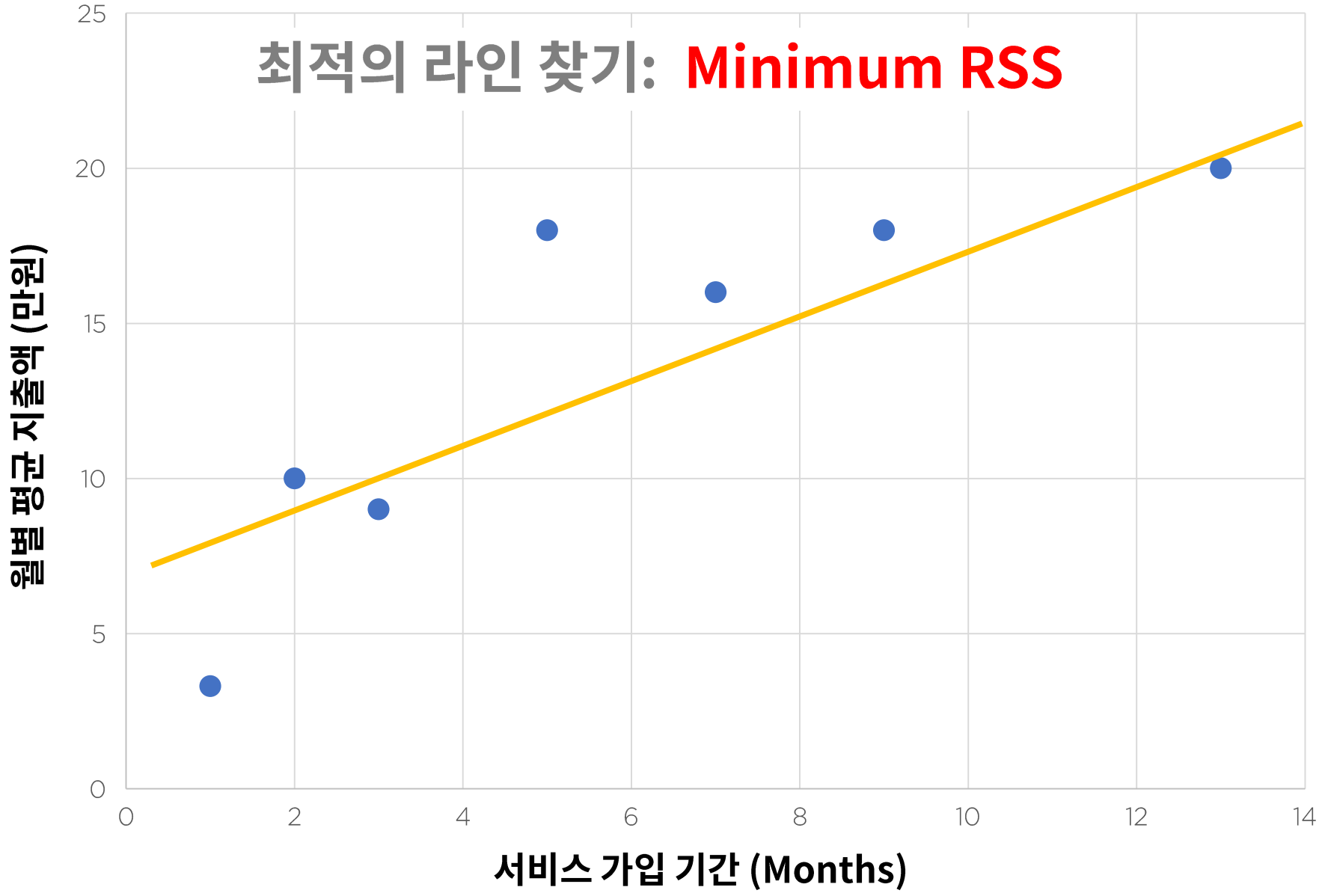












Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

R-squared, Coefficient, P-value란?

```
In [24]: lm.summary()
```

Out [24]:

OLS Regression Results

Dep. Variable:	Yearly Amount Spent	R-squared (uncentered):	0.998
Model:	OLS	Adj. R-squared (uncentered):	0.998
Method:	Least Squares	F-statistic:	4.105e+04
Date:	Mon, 31 Aug 2020	Prob (F-statistic):	0.00
Time:	23:12:08	Log-Likelihood:	-1596.0
No. Observations:	350	AIC:	3200.
Df Residuals:	346	BIC:	3215.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Avg. Session Length	12.1440	0.930	13.062	0.000	10.315	13.973
Time on App	34.1355	1.212	28.168	0.000	31.752	36.519
Time on Website	-14.3109	0.868	-16.482	0.000	-16.019	-12.603
Length of Membership	61.2897	1.250	49.025	0.000	58.831	63.749

Omnibus:	0.331	Durbin-Watson:	1.991
Prob(Omnibus):	0.848	Jarque-Bera (JB):	0.454
Skew:	-0.048	Prob(JB):	0.797
Kurtosis:	2.852	Cond. No.	55.0

[R-squared]

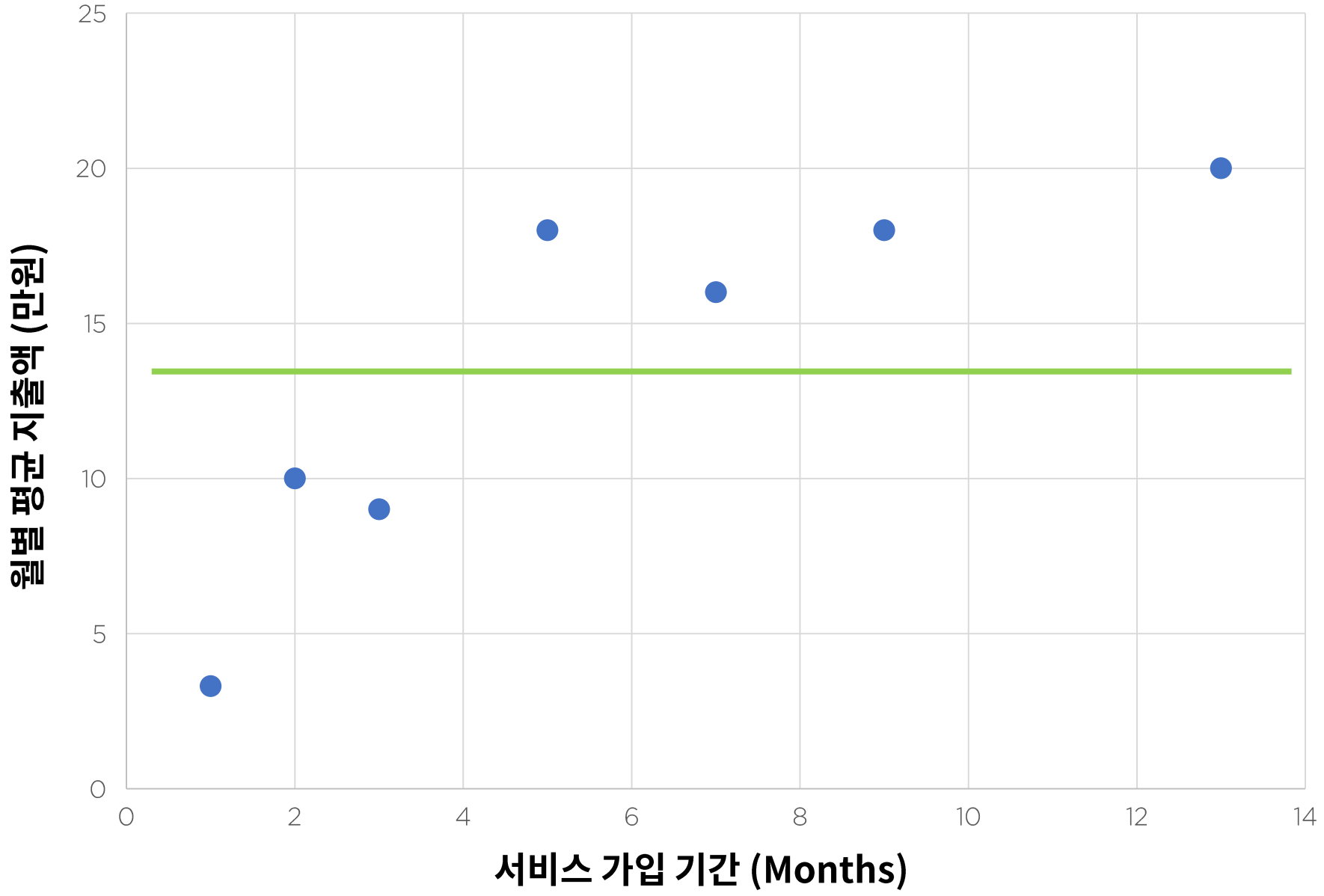
- 클수록 좋은 모델
- Adj. 가 더욱 적절한 평가 기준

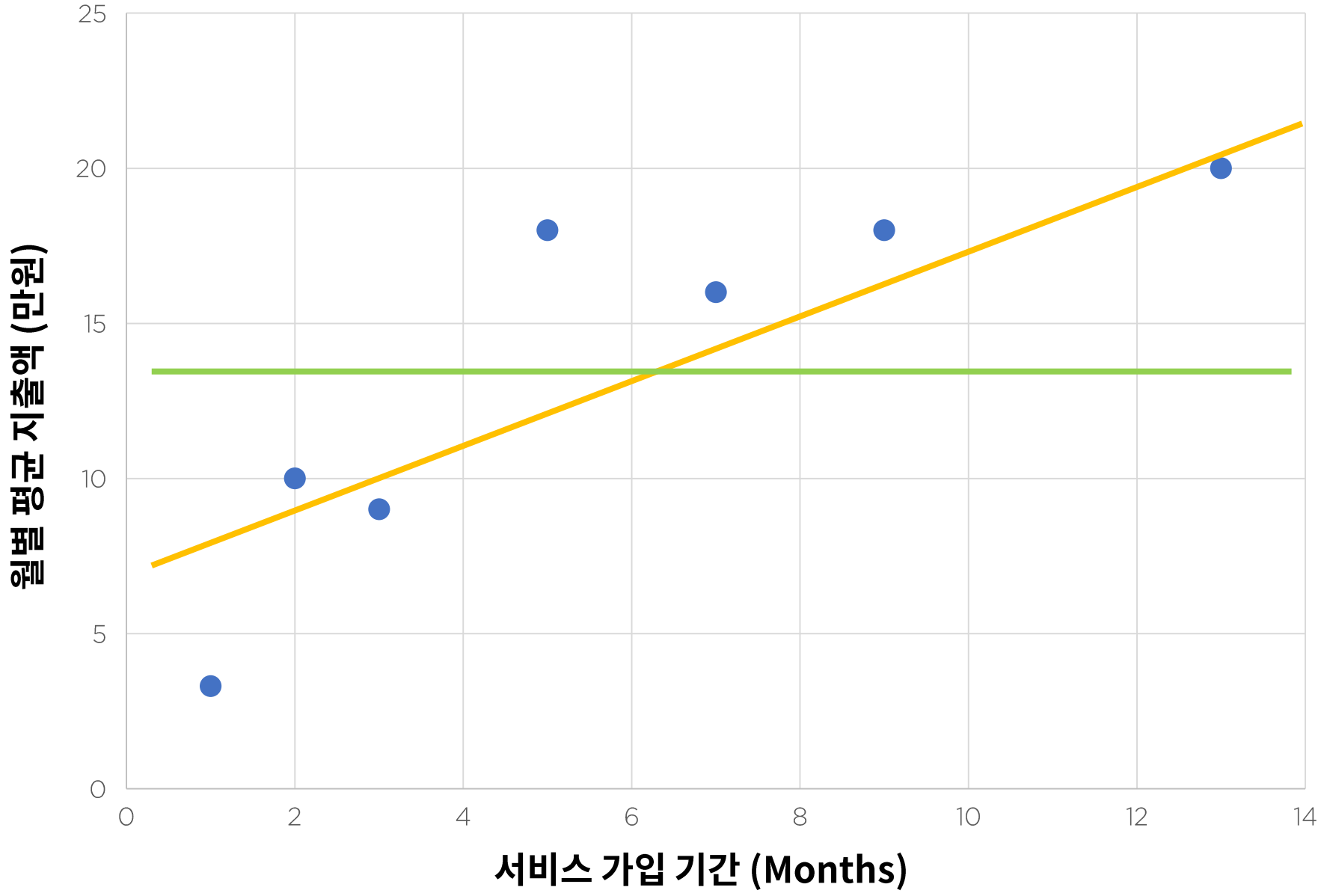
[Coef]

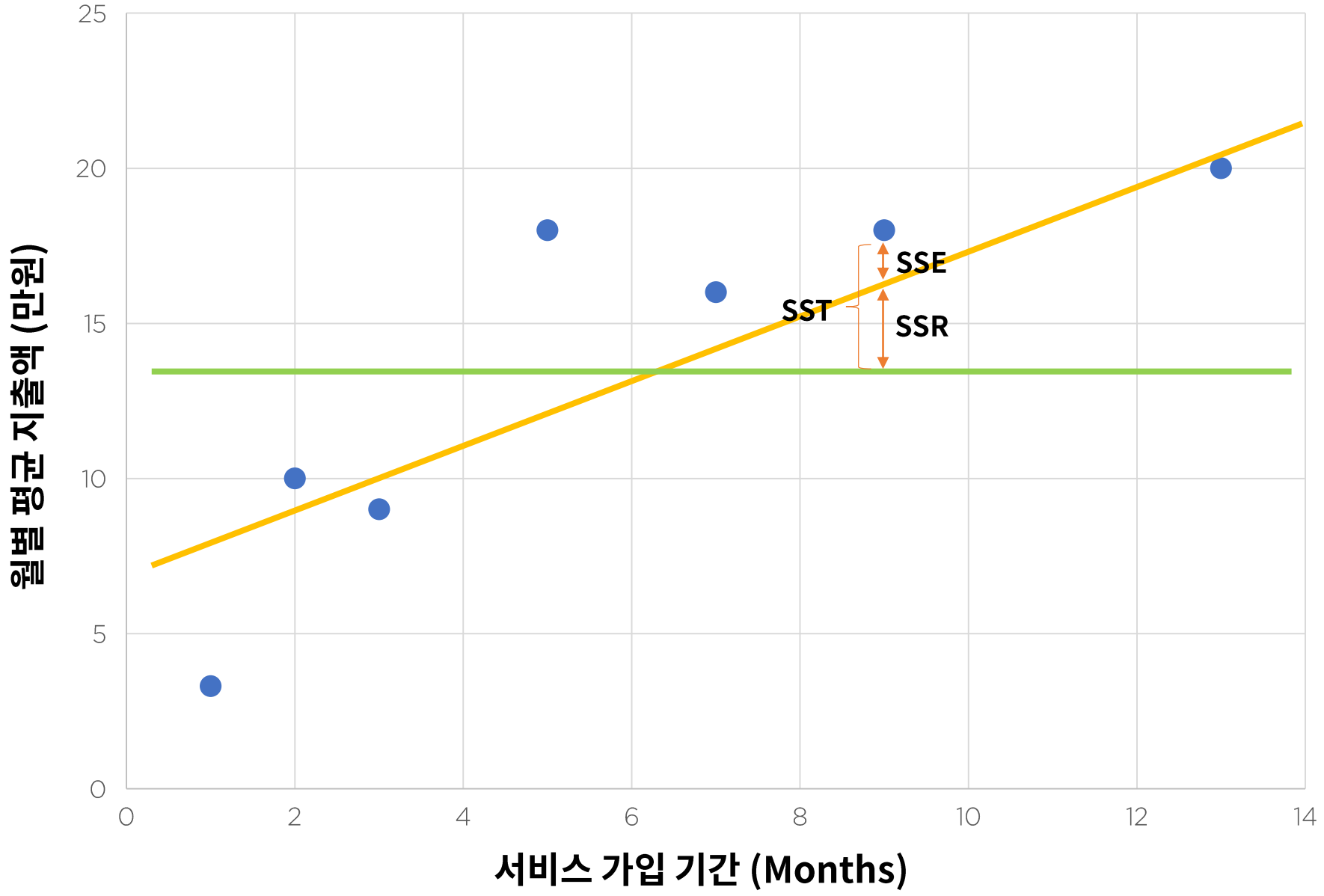
- 변수의 영향력 (강도와 방향)

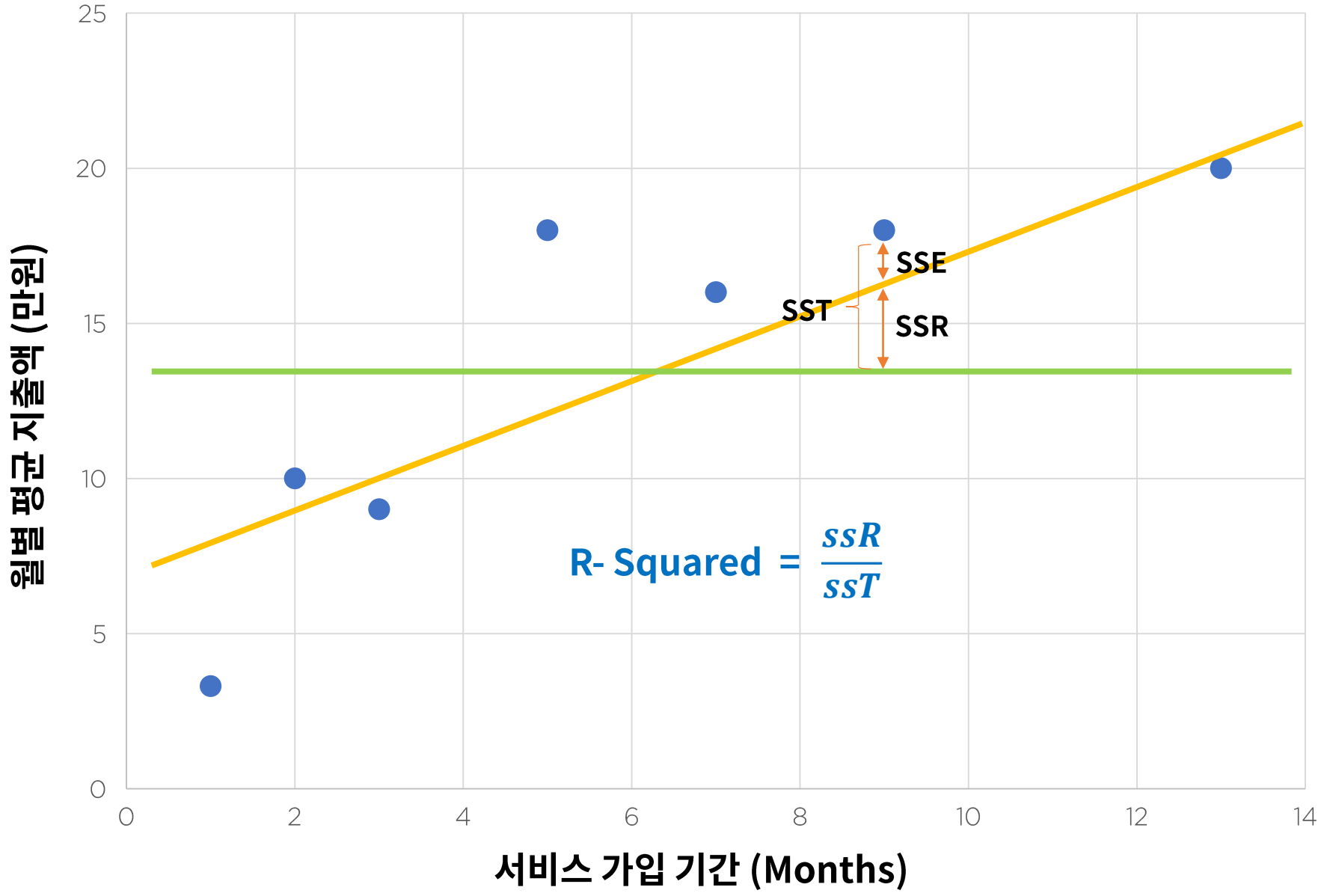
[P-value]

- 신뢰할 수 있는 결과인가?에 대한 척도
- 0.05 이하면 양호









Chapter 02. 고객별 연간 지출액 예측 (Linear Regression)

수식 만들어보기

	coef	std err	t	P>|t|	[0.025	0.975]
Avg. Session Length	12.1440	0.930	13.062	0.000	10.315	13.973
Time on App	34.1355	1.212	28.168	0.000	31.752	36.519
Time on Website	-14.3109	0.868	-16.482	0.000	-16.019	-12.603
Length of Membership	61.2897	1.250	49.025	0.000	58.831	63.749

	coef	std err	t	P>|t|	[0.025	0.975]
Avg. Session Length	12.1440	0.930	13.062	0.000	10.315	13.973
Time on App	34.1355	1.212	28.168	0.000	31.752	36.519
Time on Website	-14.3109	0.868	-16.482	0.000	-16.019	-12.603
Length of Membership	61.2897	1.250	49.025	0.000	58.831	63.749

$$y = 12.144 \times \text{Avg. Session Length} + 34.1355 \times \text{Time on App} + -14.3109 \times \text{Time on Website} + 61.2897 \times \text{Length of Membership}$$

	coef	std err	t	P>|t|	[0.025	0.975]
Avg. Session Length	12.1440	0.930	13.062	0.000	10.315	13.973
Time on App	34.1355	1.212	28.168	0.000	31.752	36.519
Time on Website	-14.3109	0.868	-16.482	0.000	-16.019	-12.603
Length of Membership	61.2897	1.250	49.025	0.000	58.831	63.749

$$y = 12.144 \times \text{Avg. Session Length} + 34.1355 \times \text{Time on App} + (-14.3109) \times \text{Time on Website} + 61.2897 \times \text{Length of Membership}$$

New Data

Avg. Session Length	33
Time on App	11
Time on Website	33
Length of Membership	2.3

444.9491