

# Evaluating Regression Models

강필성

고려대학교 산업경영공학부

pilsung\_kang@korea.ac.kr

# AGENDA

**01** Multiple Linear Regression

---

**02** Evaluating Regression Models

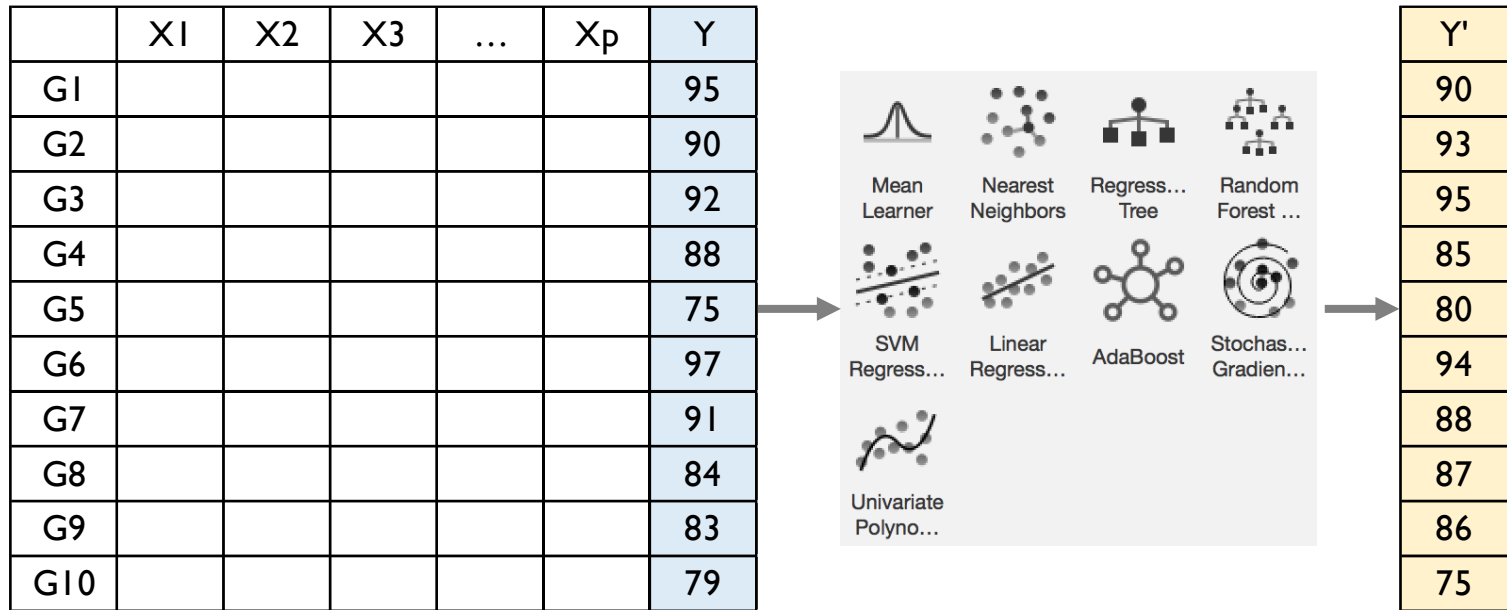
---

**03** R Exercise

---

# Evaluating Regression Models

- Example: predict a yield based on process sensors

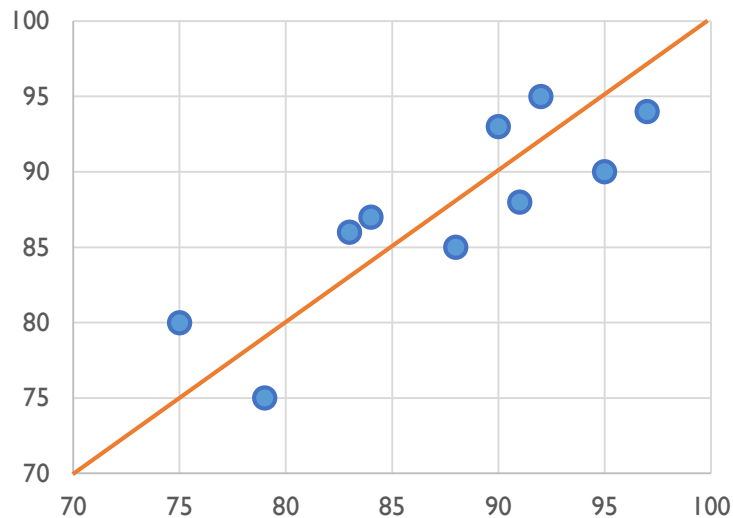


How accurate is this model?

# Evaluating Regression Models

- Performance measure I: Average Error
  - ✓ Compare the average difference between the actual and predicted y
  - ✓ Mislead to an inappropriate conclusion based on the sign effect

$$\text{Average Error} = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)$$

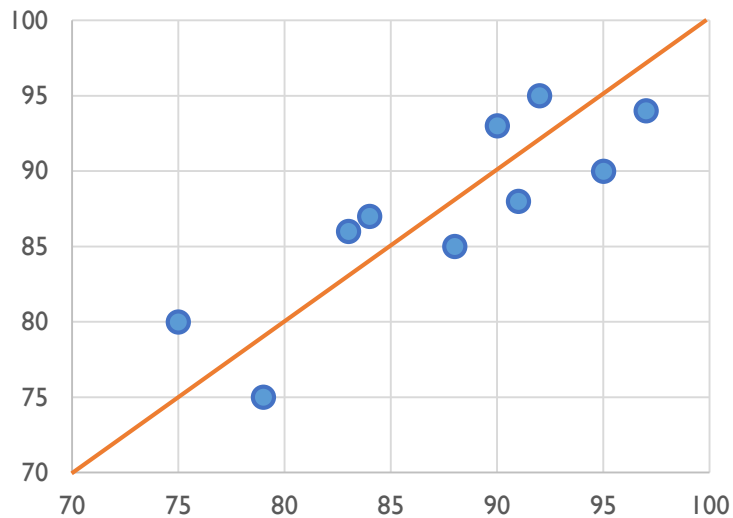


Y	Y'	Y-Y'
95	90	5
90	93	-3
92	95	-3
88	85	3
75	80	-5
97	94	3
91	88	3
84	87	-3
83	86	-3
79	75	4
Average Error		0.1

# Evaluating Regression Models

- Performance measure 2: Mean Absolute Error (MAE)
  - ✓ Compute the average of absolute value of differences between the predicted and actual  $y$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|$$



Y	Y'	Y-Y'
95	90	5
90	93	3
92	95	3
88	85	3
75	80	5
97	94	3
91	88	3
84	87	3
83	86	3
79	75	4
MAE		3.5

# Evaluating Regression Models

- Performance measure 3: Mean Absolute Percentage Error (MAPE)
  - ✓ MAE can only provide the degree of absolute difference between the predicted and actual  $y$  but cannot provide the relative difference between them
  - ✓ Ex) The MAEs of the two models below are the same ( $MAE = 1$ )

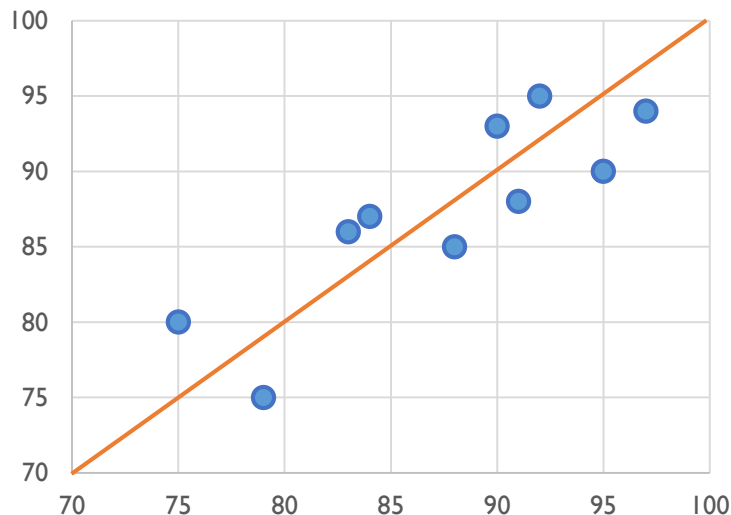
Y	Y'	Y-Y'
1	0	1
1	2	1
1	0	1
1	2	1
1	0	1
1	2	1
1	0	1
1	2	1
1	0	1
1	2	1
MAE		1

Y	Y'	Y-Y'
100	99	1
100	101	1
100	99	1
100	101	1
100	99	1
100	101	1
100	99	1
100	101	1
100	99	1
100	101	1
MAE		1

# Evaluating Regression Models

- Performance measure 3: Mean Absolute Percentage Error (MAPE)
  - ✓ Provide the relative absolute difference in terms of %
  - ✓ Commonly adopted by domains in which relative differences are more important than the absolute difference (ex: quality control in manufacturing process)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right|$$

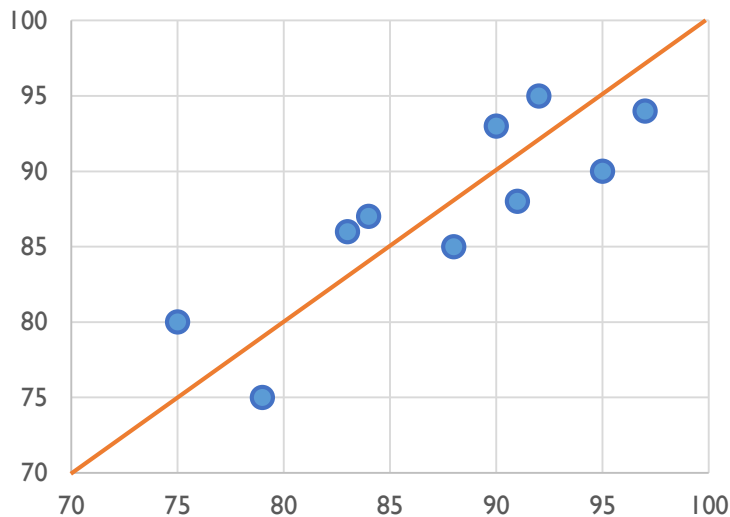


Y	Y'	Y-Y'	Y-Y' / Y
95	90	5	5.26%
90	93	3	3.33%
92	95	3	3.26%
88	85	3	3.41%
75	80	5	6.67%
97	94	3	3.09%
91	88	3	3.30%
84	87	3	3.57%
83	86	3	3.61%
79	75	4	5.06%
MAE		3.5	4.06%

# Evaluating Regression Models

- Performance measure 4 & 5: (Root) Mean Squared Error ((R)MSE)
  - ✓ Use the square instead of absolute value to resolve the effect of sign

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}$$



Y	Y'	(Y-Y') <sup>2</sup>
95	90	25
90	93	9
92	95	9
88	85	9
75	80	25
97	94	9
91	88	9
84	87	9
83	86	9
79	75	16
MSE		12.9

$$\text{RMSE} = \sqrt{12.9} = 3.59$$



