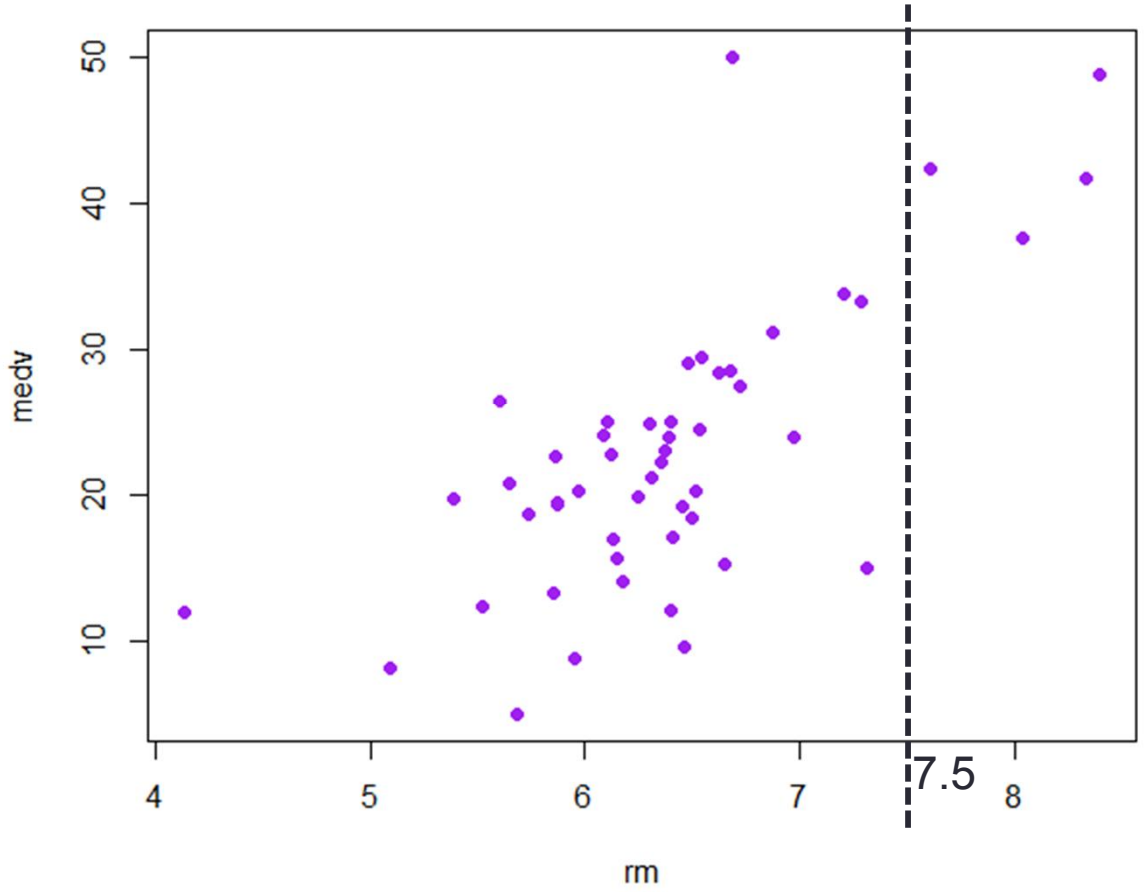
Ch.3
Linear Regression

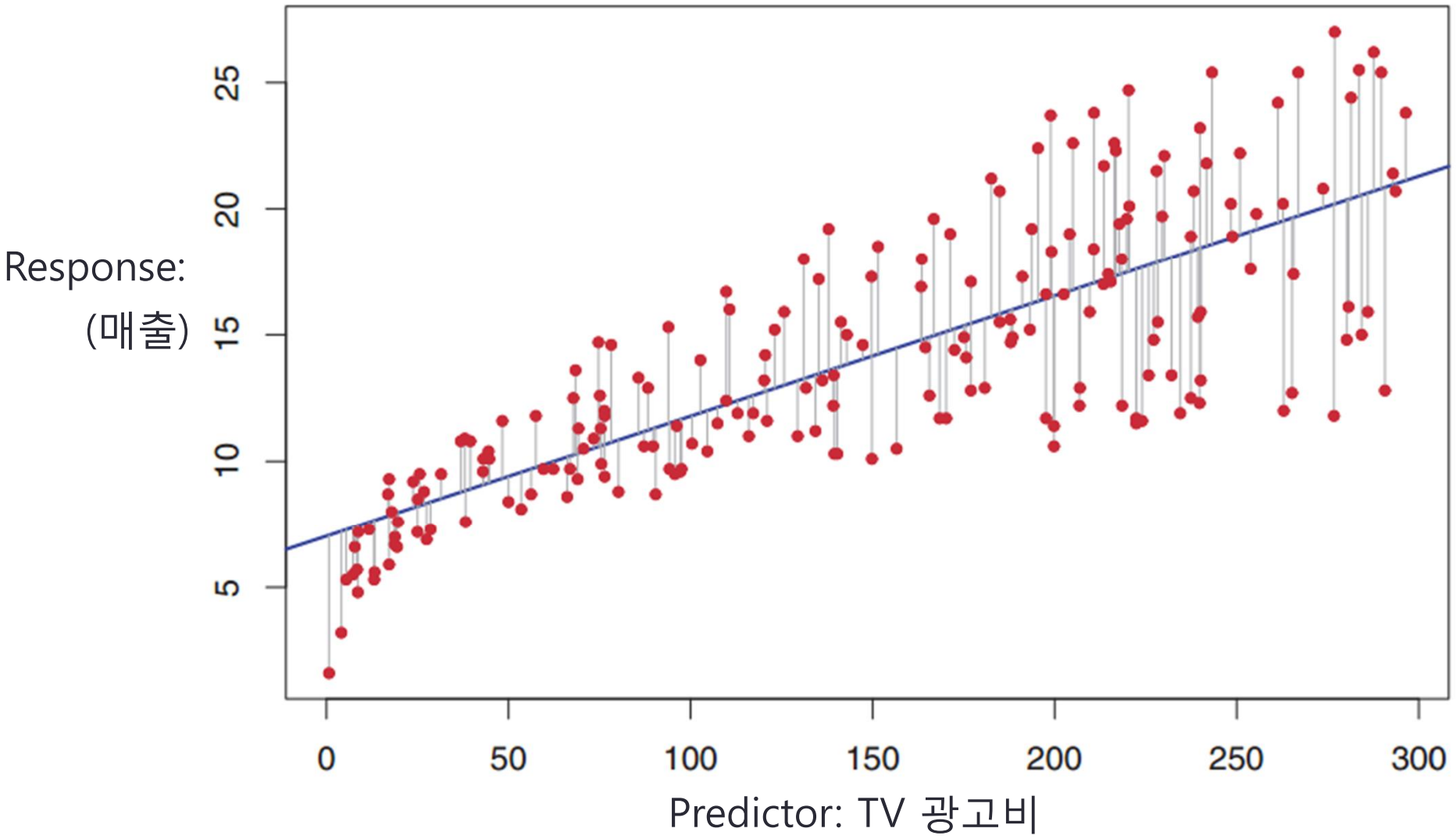
Regression : 숫자와 같이 response가 연속적이고 크기 순서가 있는 경우

Linear Regression : Regression 타입 Supervised Learning 모델 중 하나



- medv와 rm이 위와 같은 관계가 있다
- rm이 7.5 이라면 medv는 얼마일까?

Simple Linear Regression : Predictor가 1개



TV 광고비 지출에 따른 매출들이 위와 같이 200개의 • 으로 나타난다.
- 매출과 TV광고비간의 관계를 선형식으로 나타내면,

Simple Linear Regression

- labelled data (x, y)를 이용한 Supervised Learning
- 입력이 1개의 predictor로 되어 있어, coefficient는 β_0, β_1 두 개
- labelled data (x, y)를 이용해 β_0, β_1 추정치를 구하자

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- β_0, β_1 추정치 구하기 : RSS를 최소화 하는
- $\hat{\beta}_1$이 정확한가? : 샘플 (train set)이 다를 때는 어떻게 되나?

  → $\hat{\beta}_1$의 standard error 구해, 95% 신뢰도 구간 보자
- $\hat{\beta}_1$이 유효한가? : x와 y가 관계가 없을 가능성은 (즉, $\hat{\beta}_1$ = 0, Null hypothesis)

  → "t-statistic" 이 2~3 이상이면 Null hypothesis 일 가능성 작다

  → "p-value"가 0.05 이하이면 Null hypothesis 일 가능성 작다
- 모델이 데이터에 얼마나 잘 설명하나 (묘사하나)? : RSE, R^2 statistic 활용

Multiple Linear Regression Model

$$Y_{n\times 1} = X_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times 1}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon$$

* *i = 1, 2, … n* 개의 샘플들, *p=1, 2, …. p* 개의 features

- 연습 데이터 (**X**, *Y*) 를 이용해 coefficient/parameter $\beta_0, \ldots \beta_p$ 의 추정치 $\hat{\beta}_0$ ..... $\hat{\beta}_p$ 를 구함.

- y_i를 response (또는 종속변수, target, output) 라하고, x_{ij} (j=1,2, … , p)를 feature(또는 predictor, 독립변수, input) 라 함 (y_i 와 x_{ij} 는 scalar)

- β_0 is the intercept (i.e. the expected value for y if all the x's are zero), β_j (j=1,… p) 는 predictor X_j 의 coefficient(weight)

- β_j is the average increase in Y when X_j is increased by one unit while **all other X_i's are held constant.**

데이터 관련 용어

$X_{Newspaper}$

데이터프레임

TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2
8.6	2.1	1	4.8
199.8	2.6	21.2	10.6
66.1	5.8	24.2	8.6
214.7	24	4	17.4
23.8	35.1	65.9	9.2
97.5	7.6	7.2	9.7
204.1	32.9	46	19

각 행을 row 또는 instance, observation, sample, example, record라 함

각 열/Column은 Input, Output, Response, Feature, Predictor, Attribute, independent variable, Covariate 이 될 수 있음

Input, Feature, Predictor, Attribute, independent variable, Covariate 은 같은 말
- 수식에서 보통 *X*로 표현

Output, Response, Target, Outcome, Label, dependent variable 은 같은 말
- 식에서 보통 *Y*로 표현

Sales = f*(TV, Radio, Newspaper)* 로 모델을 삼으면
- *TV, Radio, Newspaper* 가 feature
- Sales가 response

Notation (ISLR의 수학 심볼 표기에 관해...)

- ISLR 수학 심볼 요약

- 모든 vector는 column vector
- n : observation/row/sample의 수
- p : feature/predictor의 수 (* column의 수와 같지 않음)
- x, X, $\mathbf{X}$: feature
  - x : scalar. x_{ij} : i 번 observation 의 j 번 feature.
    x_k : observation을 특정하지 않은 k 번 feature
  - X : 하나의 feature/predictor **vector**. X_j : j 번 (또는 'j' 명칭의) feature **vector**
  - **X** : feature vector들로 이루어진 feature **matrix** (즉, $\mathbf{X} = [\, X_1, X_2, \ldots, X_p \,]$)
- y, Y : response (* ISLR에서는 response matrix 안쓰임)
  - y : scalar. y_i : i 번 observation 의 response
  - Y : response vector. (즉, $Y = [\, y_1, y_2, \ldots, y_n \,]^T$)

Notation cont. : 대부분 일관되게 사용하나 가끔 혼란도 …

- 물론 문맥을 통해 표기가 무엇을 뜻하는 것인가 금방 알 수 있지만, 참고…

- $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p. \qquad (3.21)$$

  - 위 식은 **한 observation**의 feature tuple이 **scalar**들 $x_1, x_2, \ldots x_p$ 이고 coefficient가 위와 같을 때 response 추정치라야 말이 됨 (그런데, ISLR의 notation (1.1, p.10)에 따르면 x_i는 i번 observation의 p개 feature들의 값을 표현하는 size가 p인 **벡터** – 위의 식 (3.21)의 x_i 를 size p 벡터로 해석하면 당연히 안됨)
    → 저자가 다음 판에 notation 이슈를 고려하겠다고 답장

  - (3.21)를 벡터로 표현하면,

$$\hat{y} = \begin{bmatrix} \hat{\beta}_0, & \hat{\beta}_1, & \hat{\beta}_2, & \ldots, \hat{\beta}_p \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \beta^T X :$$

이처럼 벡터 β와 X의 dot product로 표현하는 것이 일반적

  - ISLR에서는 위의 X같이 하나의 observation의 feature들로 된 vector를 특별히 부르는 명칭이 없으나 일반적으로 pattern vector 또는 instance vector라 함(**x**, 소문자 **bold** 로 표시하기도. 따라서 $\mathbf{x}_k$: k번 observation의 feature 값들. ISLR도 이리 하면 좋겠음) . 위와 같이 1이 첨부된 것을 "*augmented* pattern vector"라고.
  - β를 coefficient vector라 하기도 하지만 많은 경우에 weight vector라 하고 *W*(또는 **w**) 로 표시하는 게 보통

표기법을 사용한 Linear System의 Vector-Matrix 식 복습

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p. \qquad (2.4)$$

⇩

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

ie. if Y can be approximated as a linear combination of $X_1, X_2, \ldots, X_p$

⇩ expand

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_p x_{1p} \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_p x_{2p} \\
y_3 &= \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33} + \cdots + \beta_p x_{3p} \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_p x_{np}
\end{aligned}$$

* β_i : 우리가 구하려는 패러미터

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \\ \\ \\ y_n \end{pmatrix} = \beta_0 + \beta_1 \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \\ \\ \\ x_{n1} \end{pmatrix} + \beta_2 \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \\ \\ \\ x_{n2} \end{pmatrix} + \beta_3 \begin{pmatrix} x_{13} \\ x_{23} \\ x_{33} \\ \\ \\ \\ x_{n3} \end{pmatrix} + \ . \ . + \beta_p \begin{pmatrix} x_{1p} \\ x_{2p} \\ x_{3p} \\ \\ \\ \\ x_{np} \end{pmatrix}$$

[augmented_pattern_vector_of_2'nd_observation]^T

$$Y = \mathbf{X} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \\ \\ \\ \beta_p \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \vdots & & & & & \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{pmatrix}$$

Training Phase:

- $\mathbf{X}_{train}$, Y_{train} 값을 이용
- RSS를 최소화하는 $\hat{\beta}$들을 구하자

Prediction Phase:

- Training을 통해 $\hat{\beta}$ 들을 구한 상황에서 새로운 $\mathbf{X}_{new}$ 가 주어지면 $\mathbf{X}_{new}\hat{\beta}$ 식으로 Y를 추정한다.

Regression Coefficient 추정

여기에서 X, Y는 scalar

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_p X_{ip}$$ 에서 $\hat{\beta}_0 \ldots \hat{\beta}_p$ 구하기 :

sum, square, residual

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

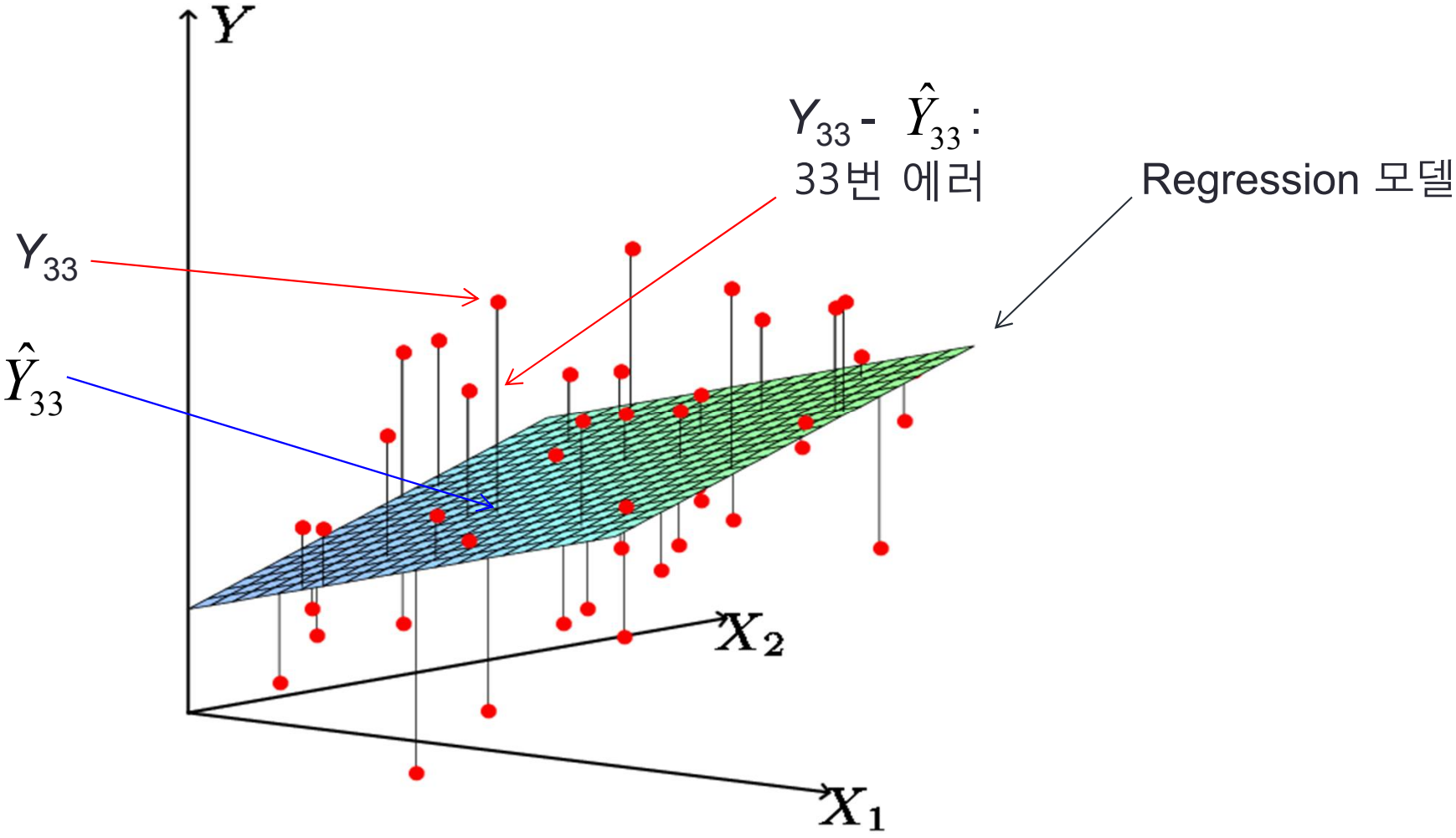
RSS : Residual Sum of Squares :
(실제 Y값과 추정한 Y 값과의 오차)² 들의 합

$$= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2$$

- (Ordinary) Least Squares Method:
  - RSS를 최소화시키는 $\hat{\beta}_0 \ldots \hat{\beta}_p$ 를 구하자

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

RSS : (에러들의 길이)²의 합



Relationship between population and least squares lines

Population

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Least Squares Estimation

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p$$

- $\beta_0 , \ldots . \beta_p$ 를 실제에선 얻지 못하기에 Least Squares Estimation으로 그대신 $\hat{\beta}_0 \ldots . \hat{\beta}_p$ 를 구해 사용
- $\hat{\beta}_0 \ldots \ldots \hat{\beta}_p$ 가 추정치이기에 $\hat{Y}_i$ 또한 추정치. 즉 실제와 오차 (residual/error) 가 생김.

모델 정확성 평가 : R^2 statistic

- R^2 measures the proportion of variability in response Y that can be explained by using predictors X.

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{RSS}{\sum (Y_i - \bar{Y})^2}$$

TSS : Total Variance in Y

- **R^2 은 0 과 1 사이의 값을 갖는다.**
- A number near 0 indicates that the regression did not explain much of the variability in the response; this might occur because the linear model is wrong, or the inherent error σ^2 is high, or both.
- An R^2 statistic that is close to 1 indicates that a large proportion of the variability in the response has been explained by the regression (즉, regression 모델이 유효)
- **Multiple linear regression**에서 R^2은 predictor를 늘릴 때마다 증가한다. 만약 새로운 predictor가 R^2 를 조금밖에 (가령 0.01이하) 증가시키지 못하면 그 predictor를 추가함이 적절한 것이 아닐 수 있다.
- R^2 값이 작다고 (예, $R^2 = 0.1$) 모델이 의미없다는 말은 아니다

Some Important Questions

1. *Is at least one of the predictors $X_1, X_2, \ldots, X_p$ useful in predicting the response?*

2. *Do all the predictors help to explain Y, or is only a subset of the predictors useful?*

3. *How well does the model fit the data?*

4. *Given a set of predictor values, what response value should we predict, and how accurate is our prediction?*

1. Predictor와 Response간 관계가 있나?

1) **p** 개의 Predictor중 하나라도 response와 관계가 있는 가를 따져본다

2) 우선 **F-statistic**을 본다.

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} > 1$$

1) **response**가 그 어떤 predictor와도 관계가 없으면 F-statistic →1이며, 관계가 있으면 1보다 커진다

2) **F-statistic**의 해석은 n(observation 개수)과 p(predictor 개수)를 고려해야 함. n이 크면 F-statistic이 1에 가까워도 null hypothesis를 거부할 수 있으나, n이 작으면 F-statistic이 1보다 충분히 커야 커야 함

3) F-statistic의 p-value가 충분히 작으면 ($\ll 0.05$), 최소한 predictor중 하나라도 response와 관계가 있다고 생각할 수 있다. F-statistic의 p-value 가 0.05보다 훨씬 크면 이 모델을 더 볼 이유가 없다.
* t-statistic이 0에서 멀어질수록, null hypothesis를 거부하기 쉬움

4) 구체적으로 어떤 predictor가 의미가 있는가를 보려면 개별 predictor의 t-statistic/p-value를 본다. 개별 predictor의 p-value는 그 predictor가 model에 추가되어 response에 추가로 영향을 미치는 정도(partial effect)를 나타냄. 보통 $p<0.05$ 이고 $n>p$ 이면 그 predictor에 대한 null hypothesis를 거부할 수 있다.

- F-statistic과 그 p-value를 보지 않고 개별 predictor의 p-value만으로 모델/predictor의 유효성을 판단함은 큰 문제가 있을 수 있다. 왜냐하면, predictor의 개수가 많을 때 순전히 "운"으로도 일부 predictor가 response와 관계가 있는 것처럼 비칠 수 있고 predictor의 p-value가 0.05 이하 될 수 있다. Hastie&Tibshirani에 의하면 100개의 predictor가 있으면 실제로 predictor와 response가 아무 관계가 없을 때에도 5개 정도는 p-value < 0.05 라 함

- p > n 이면, observation 수보다 많은 수의 coefficient 들을 구해야 한다는 말로, 이 때는 전체 predictor를 사용할 경우 regression 자체가 불가능하다.

2. 중요한 변수/Predictor는 어떤 것일까?

1) 앞에서 언급한 방식으로 F-statistic p-value를 본 결과, 최소한 변수 하나는 response와 관련이 있다 판단되면,

2) 어떤 변수가 중요한 변수인가 판별할 때,

- **Forward Selection :** We begin with the null model - a model that contains an intercept but no predictors. We then fit ***p simple*** linear regressions and add to the null model the variable that results in the lowest RSS(higest R^2) for the new two-variable model. This approach is continued until some stopping rule is satisfied

- **Backward Selection :** We start with all variables in the model, and remove the variable with the largest p-value. The new (p−1) variable model is fit, and the variable with the largest p-value is removed. This procedure continues until a stopping rule is reached. For instance, we may stop when all remaining variables have a p-value below some threshold.

- **Mixed Selection :** This is a combination of forward and backward selection. We start as with forward selection. Forward selection으로 진행하다, 어느 순간 새로운 변수를 넣은 후 모델에 있는 변수 중 하나의 p-value가 설정값보다 커지면, 그 변수를 모델에서 제거. 이렇게 계속 진행하면 모델에 포함되지 않은 모든 변수는, 그 것을 모델에 넣으면 모델의 어떤 변수의 p-value를 설정치보다 높이게 되는 것들만 남는다.

➢ Backward Selection은 p > n이면 쓸 수 없다. Forward Selection은 가능

3. 모델의 성능 – 얼마나 모델이 데이터를 잘 설명하나?

1) 가장 많이 사용하는 수치로 RSE (Residual Standard Error)와 R^2가 있다

$$\mathrm{RSE} = \sqrt{\frac{1}{n-p-1}\mathrm{RSS}}$$

n > p +1 시 의미

2) RSE는 RSS나 p가 커지면 같이 커지고, n이 커지면 작아진다. 또, RSS는 Y 단위를 따르므로 RSE도 Y 단위를 따른다

3) R^2은 Y단위에 관계없이 모델이 얼마나 Y의 variance를 잘 설명하나를 나타내며 0 ~ 1 값을 갖는다. R^2 값이 1에 가까워 질수록 좋은 모델. 허나 R^2 값은 새로운 predictor를 모델에 포함할 때마다 계속 증가를 하므로, 변수 선정에 주의를 해야 함

$R^2 = \mathrm{Cor}(Y, \hat{Y})^2$

모델이 바뀌면서 R^2 값이 증가 또는 감소하나를 보고, 증가하면 증가율을 고려한다.

새로운 변수를 모델에 포함했는데 R^2가 조금밖에 증가하지 않으면 그 변수 제외 고려.

4) 그래프를 그려 보아 모델의 유효성, 부족한 면을 판단

4. 모델의 예측 성능은 어떤가?

1) We can use Confidence Interval to quantify the uncertainty surrounding the average response

2) We can use Prediction Interval to quantify the uncertainty surrounding the particular response

Other Considerations in the Regression Model

❖ Qualitative(카테고리형) Predictor를 어떻게 다루나?

- 카테고리형 변수 : R에서는 factor 라 함
- 남/여 **두가지 카테고리(2 레벨 in R)**를 갖는 gender변수
- gender의 두 카테고리를 숫자로 나타낼 수 있는 indicator/dummy 변수 1개를 만든다

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male,} \end{cases}$$ 남자를 baseline 삼음

- regression equation에 x_i를 사용하면,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

- R은 카테고리형 변수가 factor이면 자동적으로 indicator변수로 만들고, 이를 regression에 적용함!

3개 이상의 레벨을 지닌 Qualitative Predictor는?

- k 개의 카테고리(레벨)을 갖는 카테고리형 변수를 나타내기 위해서는 k-1 개의 indicator 변수가 필요
- 아시아인, 백인, 흑인 3 인종을 나타내는 ethnicity 변수를 regression에 포함하려면; 2개의 indicator 변수를 준비

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person is Asian} \\ 0 & \text{if } i\text{th person is not Asian,} \end{cases}$$

$$x_{i2} = \begin{cases} 1 & \text{if } i\text{th person is Caucasian} \\ 0 & \text{if } i\text{th person is not Caucasian} \end{cases}$$

- 위 두개의 dummy 변수를 regression에 사용하면;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is African American} \end{cases}$$

* β_1, β_p : 양수, 또는 음수일 수 있음

예) 3개 이상의 레벨을 지닌 Qualitative Predictor 경우

	Coefficient	Std. error	t-statistic	p-value
Intercept	β_0 531.00	46.32	11.464	< 0.0001
ethnicity[Asian]	β_1 −18.69	65.02	−0.287	0.7740
ethnicity[Caucasian]	β_2 −12.50	56.68	−0.221	0.8260

TABLE 3.8. *Least squares coefficient estimates associated with the regression of* balance *onto* ethnicity *in the* Credit *data set.*

F-statistic p-value: 0.96

$$\text{balance}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

- 흑인일 경우 : x_{i1}, x_{i2} 모두 0. 따라서 balance = 531
- 아시안인 : x_{i1} = 1, x_{i2} = 0. 따라서 balance = 531 – 18.69
- 백인 : x_{i1} = 0, x_{i2} = 1. 따라서 balance = 531 – 12.50

* 하지만 F-statistic의 p-value가 매우 크다. 따라서 이 경우 null hypothesis 를 거부할 수 없어 인종과 balance간에 관계가 없다 판단함이 맞다

❖ Interaction Effect

- 한 predictor x_k 가 response Y에 미치는 영향이 다른 predictor 들의 값과 무관하다는 predictor의 “additive” 특성을 따르는 것이 맞지 않을 때
- Interaction Effect = 시너지 효과
- predictor x_k의 값이 변하면, 해당 β_k 만큼만 response Y에 영향을 주는 것이 아니라, x_k와 interaction(시너지) 관계에 있는 다른 predictor의 coefficient에 영향을 미쳐 결과적으로 response Y에 끼치는 영향이 단순한 additive 형태가 아닐 때
- Predictor간 인터액션 관계를 표현하려면, “interaction term” 이라 부르는 predictor를 생성하며, “**predictor간 곱”**으로 만든다
- R/Python은 간단한 formula 로 다양한 Interaction 구현

예) Interaction in advertising

$$Sales = \beta_0 + \beta_1 \times TV + \beta_2 \times Radio + \beta_3 \times (TV \times Radio)$$

Interaction Term : TV와 라디오가 시너지 효과

$$Sales = \beta_0 + (\beta_1 + \beta_3 \times Radio) \times TV + \beta_2 \times Radio$$

- **TV**를 통해 **$1** 을 더 쓰면, **(0.019 +0.001xRadio)** 만큼 평균 매출이 늘어난다. **"0.001xRadio"** 이 있으므로 만약 **Radio**를 통한 광고가 있으면 **TV** 광고 효과가 더욱 증대된다

Parameter Estimates

	Term	Estimate	Std Error	t Ratio	Prob>|t|
β_0	Intercept	6.7502202	0.247871	27.23	<.0001 *
β_1	TV	0.0191011	0.001504	12.70	<.0001 *
β_2	Radio	0.0288603	0.008905	3.24	0.0014 *
β_3	TV*Radio	0.0010865	5.242e-5	20.73	<.0001 *

Leave the main effect Predictor

It is sometimes the case that an interaction term has a very small p-value, but the associated main effects (in this case,TV and radio) do not. The hierarchical principle states that if we include an interaction in a model, we should also include the main effects, even if the p-values associated with their coefficients are not significant. In other words, if the interaction between X1 and X2 seems important, then we should include both X1 and X2 in the model even if their coefficient estimates have large p-values.

선형회귀분석 의 가정과, 그 가정이 성립하는 경우 또는 그렇지 않은 경우를 잘 이해하여 회귀분석의 결과를 활용

There are a number of possible problems when fitting the linear regression model to a particular data set

1. Non-linearity of the data
2. Correlation of error terms
3. Non-constant variance of error terms (homoscedasticity 가정이 어긋남 - Heteroscedasticity)
4. Outliers
5. High leverage points (매우 특이한 값의 predictor)
6. Collinearity

ISLR 3.3.3 꼭 읽어 볼 것을 권고

formula = "Sales ~ Advertising+Age+CompPrice+Education+Income+Population+Price+ShelveLoc+US+Urban - Population – Education + ShelveLoc:Advertising + Income:Advertising", data = Carseats

Dep. Variable:	Sales	**R-squared:**	0.876
Model:	OLS	**Adj. R-squared:**	0.872
Method:	Least Squares	**F-statistic:**	227.1
Date:	Tue, 18 Oct 2016	**Prob (F-statistic):**	8.23e-167
Time:	17:34:46	**Log-Likelihood:**	-565.42
No. Observations:	400	**AIC:**	1157.
Df Residuals:	387	**BIC:**	1209.
Df Model:	12		
Covariance Type:	nonrobust		

Indicator variable

Interaction Terms

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	5.6339	0.519	10.849	0.000	4.613 6.655
ShelveLoc[T.Good]	4.8258	0.217	22.206	0.000	4.399 5.253
ShelveLoc[T.Medium]	2.0352	0.174	11.680	0.000	1.693 2.378
US[T.Yes]	-0.1546	0.146	-1.056	0.292	-0.442 0.133
Urban[T.Yes]	0.1299	0.112	1.160	0.247	-0.090 0.350
Advertising	0.0762	0.027	2.796	0.005	0.023 0.130
ShelveLoc[T.Good]:Advertising	0.0043	0.023	0.188	0.851	-0.041 0.049
ShelveLoc[T.Medium]:Advertising	-0.0114	0.019	-0.594	0.553	-0.049 0.026
Age	-0.0458	0.003	-14.406	0.000	-0.052 -0.040
CompPrice	0.0931	0.004	22.634	0.000	0.085 0.101
Income	0.0109	0.003	4.207	0.000	0.006 0.016
Price	-0.0952	0.003	-35.929	0.000	-0.100 -0.090
Income:Advertising	0.0008	0.000	2.701	0.007	0.000 0.001