통계적 예측모형

서울시립대 통계학과 이용희

2024-03-20

목차

Pr	eface		1				
1.	선형	회귀모형의 소개	3				
	1.1.	예제-단순 회귀모형	3				
	1.2.	선형 회귀모형	6				
	1.3.	최소제곱법	7				
	1.4.	결정계수	15				
	1.5.	중회귀 모형	18				
		1.5.1. 최소제곱추정	20				
	1.6.	예제-중회귀모형	21				
	1.7.	최소제곱 추정량의 분포	22				
	1.8.	가우스-마코브 정리	23				
	1.9.	최대가능도 추정	25				
2.	선형	회귀에서의 추론	29				
	2.1.	제곱합의 분포	29				
		2.1.1. 잔차제곱합의 분포	29				
		2.1.2. 회귀제곱합의 분포	30				
		2.1.3. 잔차제곱합과 회귀제곱합의 독립	32				
		2.1.4. 총제곱합의 분포	32				
	2.2.	모분산의 추정	32				
	2.3.	최소제곱 추정량의 성질	33				
	2.4.	모형의 적합도 검정과 분산분석	33				
3.	모형	모형의 비교 					
	3.1.	직교하는 설명 변수	35				
	3.2.	설명변수의 추가	39				
	3.3.	부분 F-검정과 가능도비 검정	45				
4.	모형.	의 진단	49				
	_		49				
		4.1.1. 가중 최소제곱법	50				
	4.2.	변수변환	52				
		4.2.1. 지수모형과 멱함수: 로그변환	52				
		4.2.2. 쌍곡선과 역변환	54				
		4.2.3 Roy-Coy 변화	56				

	중공선성	57								
	4.	3.1. 독립변수간의 상관계수	57							
	4.	3.2. X ^t X 의 고유값(Eigenvalues)	57							
	4.	3.3. 조건지수 (condition number)	59							
	4.	3.4. 분산팽창계수 (Variance Inflation Factor ;VIF)	59							
Re	eference	s ·	60							
Αŗ	pendic	es e	51							
Α.	행렬의 기	기초	61							
	A.1. 벡	터와 행렬	61							
	A.2. 두	행렬의 덧셈	61							
	A.3. 스	칼라곱	62							
	A.4. 벡	터와 행렬의 곱셈	62							
	A	4.1. 행과 열의 내적	62							
	A	4.2. 열벡터의 선형조합	63							
	A.5. 행	렬의 전치	63							
	A.6. 행	렬의 곱셈	63							
	A.7. 단	위벡터와 항등행렬	65							
	A.8. 대	각합	66							
	A.9. 행	렬식	67							
	A.10.직	교행렬	67							
В.	벡터공긴	벡터공간								
	B.1. 벡	터공간의 정의와 의미	69							
	B.2. 벡	터의 선형독립	71							
			72							
		ㅇ _ 렬의 계수								
	B.5. 생		73							
			74							
	B.7. 행	르의 열공간과 행공간	74							
	B.8. 두	벡터의 사영	75							
	B.9. 최	소제곱법과 사영	76							
C	고유강고	· 고유벡터	78							
٠.			78							
	-		78							
			78							
			79							
		" -	79							
			83							

D.	벡터	미분	84				
	D.1.	용어	84				
	D.2.	벡터 미분의 표기법	84				
	D.3.	함성함수의 미분법	86				
	D.4.	두 벡터 내적의 미분	87				
		D.4.1. 상수벡터와 변수벡터의 내적	87				
		D.4.2. 상수벡터와 함수벡터의 내적	87				
		D.4.3. 함수벡터와 함수벡터의 내적	88				
	D.5.	벡터 미분의 응용	88				
		D.5.1. 선형사상의 미분상	88				
Ε.	다변	량 확률변수의 성질	90				
	E.1.	일변량분포	90				
	E.2.	확률벡터와 분포	90				
	E.3.	다변량 정규분포	92				
	E.4.	표준정규분포로의 변환	94				
	E.5.	예제	95				
F.	이차형식과 제곱합의 분포 9						
	F.1.	이차형식	96				
	F.2.	대칭행렬의 대각화	97				
	F.3.	멱등행렬	98				
	F.4.	이차형식의 분포	98				
		F.4.1. 카이제곱 분포	98				
		F.4.2. 비중심 카이제곱 분포	99				
		F.4.3. 이차형식의 분포	99				
		F.4.4. 이차형식의 독립	100				
		F.4.5. 이차형식의 차이	100				
	F 5	코그라이 저리	101				

Preface

이 책은 통계적 예측모형에 대한 교재이며 일반 선형모형을 포함하여 예측에 사용되는 기본적인 통계 모형에 대한 이론을 최대 가능도 추정법의 관점에서 설명합니다. 또한 실제 예제를 통한 실습, 모형을 적합하는 계산방법과 연관된 행렬이론에 대하여 다루고자 합니다.

i 노트

- 이 책에서 사용된 기호, 표기법, 프로그램의 규칙과 쓰임은 다음과 같습니다.
 - 스칼라(scalar)와 일변량 확률변수는 일반적으로 보통 글씨체의 소문자로 표기한다. 특별한 이유가 있는 경우 대문자로 표시할 것이다.
 - 벡터, 행렬, 다변량 확률벡터는 굵은 글씨체로 표기한다.
 - 통계 프로그램은 R을 이용하였다. 각 예제에 사용된 R 프로그램은 코드 상자를 열면 나타난다.

강의의 부교재는 강근석 와/과 유형조 (2016) 을 사용한다.

이 교과서에서 이용하는 R 패키지는 다음과 같다.

```
library(here)
                       # file pathways
library(tidyverse)
                       # data management, summary, and visualization
library(MASS)
library(knitr)
library(kableExtra)
library(agricolae)
library(emmeans)
library(plotly)
library(plot3D)
# 아래 3 문장은 한글을 포함한 ggplot 그림이 포함된 HTML, PDF로 만드는 경우 사용
library(showtext)
font_add_google("Nanum Pen Script", "gl")
showtext_auto()
#강의 부교재 자료를 포함한 패키지 설치
install.packages("remotes")
```

Preface

remotes::install_github("regbook/regbook")

library(regbook)

1.1. 예제-단순 회귀모형

보기 1.1 (자동차의 제동거리). 자동차가 달리는 속도(speed,단위는 mph; mile per hour)와 제동거리(dist, 단위는 ft;feet)의 관계를 알아보기 위하여 50대의 자동차로 실험한 결과의 자료 cars 는 다음과 같다(처음 10개의 자료만 보여준다). 자료는 R의 data.frame 형식으로 저장되어 있다.

아래 자료를 보면 실험에서 2대의 자동차는 7 mph 로 달리다가 브레이크를 밟고 정지하는 경우 각각 4, 22 feet 의 제동거리가 필요한 것으로 나타났다. 또한 3대의 는 10 mph 로 달리다가 각각 18, 26, 34 feet 의 제동거리가 필요한 것으로 나타났다.

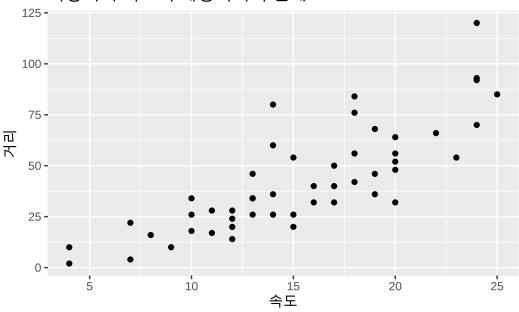
cars %>% head(n=10)

	speed	dist
1	4	2
2	4	10
3	7	4
4	7	22
5	8	16
6	9	10
7	10	18
8	10	26
9	10	34
10	11	17

자동차의 속도와 제동거리에 대한 산포도는 아래와 같다.

```
ggplot(cars, aes(x=speed, y=dist)) + geom_point() + labs(x = "속도", y = "거리") + labs(title="자동차의 속도와 제동거리의 관계")
```

자동차의 속도와 제동거리의 관계



위와 같은 자료를 이용하여 자동차의 속도가 주어졌을 경우 제동거리를 예측하려고 한다면 어떤 방법을 사용해야 할까?

보기 1.2 (아파트 판매가격). 다음 살펴볼 자료는 2019년 거래된 서울 아파트의 실거래 데이터 중 4개의 구(동대 문구, 서초구, 관악구, 노원구)에서 거래된 아파트 중 1000개의 아파트를 임의로 추출한 자료이다.

apart_2019 <- read.csv(here("data", "seoul_apartment_2019_sample.csv"), header = T)
head(apart_2019,10)</pre>

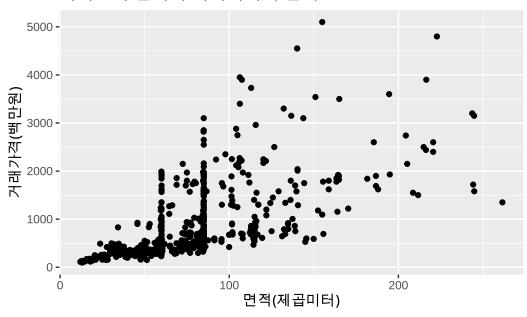
gu year area price

- 1 관악구 1974 65.09 450
- 2 관악구 1978 56.86 276
- 3 관악구 1982 91.24 599
- 4 관악구 1982 91.24 560
- 5 관악구 1984 60.27 425
- 6 관악구 1984 60.27 420
- 7 관악구 1985 38.92 217
- 8 관악구 1988 61.74 485
- 9 관악구 1991 71.90 328
- 10 관악구 1991 84.44 438

아파트의 면적(area;제곱미터)에 따른 거래가격(price;백만원)의 변화는 다음 그림과 같다.

ggplot(apart_2019, aes(x=area, y=price)) + geom_point() + labs(x = "면적(제곱미터)", y = "거래가격이 관계")

아파트의 면적과 거래가격의 관계



만약 아파트의 면적(x)과 거래가격(y) 대신 각각의 로그값 $(\log(x), \log(y))$ 을 시용하면 다음과 같은 산포도가 나타난다.

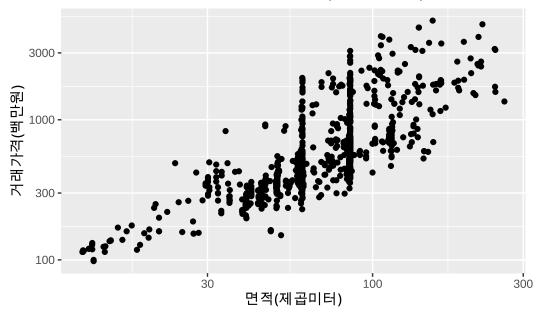
```
# log scale for y

ggplot(apart_2019, aes(x=area, y=price)) + geom_point() + labs(x = "면적(제곱미터)", y = "거래가격이 scale_y_log10() +

scale_x_log10() +

labs(title = "아파트의 면적과 거래가격의 관계(로그스케일)")
```

아파트의 면적과 거래가격의 관계 (로그스케일)



1.2. 선형 회귀모형

회귀모형(regression model)는 변수들의 함수적 관계를 분석하는 통계적 방법이다. 일반적으로 한 개 또는 여러 개의 설명변수들(explanatory variables, \mathbf{x})이 관심있는 반응변수(response variable, \mathbf{y})에 어떤 형태로 영향을 미치는지에 파악하고 설명변수와 반응변수의 함수 관계를 통계적으로 추론하는 것이 회귀분석의 목적이다.

위에서 살펴본 두 예제에서 자동차의 속도(x)가 증가하면 제동거리(y)가 증가하는 경향이 있다는 것을 알 수 있으며, 아파트의 면적(x)과 거래가격(y)도 유사한 관계임을 알 수 있다.

이러한 두 변수의 관계를 다음과 같은 반응변수 y 와 설명변수의 선형 에측식(linear predictor)으로 나타내어 보자. 이러한 관계는 반응변수의 값의 변화를 근사적으로 설명변수의 선형식으로 예측할 수 있다는 의미이다.

$$y \approx \beta_0 + \beta_1 x$$

위와 같은 근사적인 관계를 더 구체화하여 다음과 같이 반응변수의 평균값이 설명변수의 선형식으로 나타나는 것을 가정할 수 있으며 이를 선형 회귀모형(linear regression model)이라고 한다.이

$$E(y|x) = \beta_0 + \beta_1 x \tag{1.1}$$

식 1.1 은 반응변수 y의 평균이 설명변수 x 의 선형 예측식으로 나타나는 관계를 가정한 것이며 절편 β_0 와 기울기 β_1 는 회귀계수(regression coefficient)라고 부르는 모수(parameter)로서 추정해야 한다.

특별히 하나의 설명변수를 사용하는 회귀 모형을 단순선형 회귀모형(simple linear regression model)이라고 한다.

위에서 본 두 예제와 같이 n 개의 자료 $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ 을 독립적으로 추출하였다면 자료의 생성 과정을 다음과 같은 단순선형 회귀모형으로 나타낼 수 있다. 반응변수 y_i 는 설명변수 x_i 의 선형함수로 표현된 선형예측식 식 1.1 과 임의의 오차항 (random error) e_i 의 합으로 나타내어진다고 가정하자.

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \dots, n$$
 (1.2)

여기서 오차항 e_i 는 평균이 0이고 분산이 σ^2 인 임의의 확률분포를 따르며 서로 독립이라고 가정한다.

$$E(e_i) = 0$$
, $V(e_i) = \sigma^2$ $i = 1, 2, ..., n$

오차항의 분산 σ^2 도 추정해야할 모수(parameter)이다.

1.3. 최소제곱법

앞에서 언급한 것과 같이 선형회귀모형 식 1.2 에서 모수 β_0 와 β_1 를 회귀계수라고 하며 자료를 이용하여 추정해야 한다. n개의 자료를 이용하여 회귀계수 β_0 와 β_1 를 추정하려고 할 때 사용할 수 있는 방법들 중에서 가장 쉽고 유용한 방법은 최소제곱법(least square method)이다.

회귀모형 식 1.2 에서 β_0 와 β_1 의 값이 주어졌다면 설명변수 x_i 에서 반응변수의 관측값 y_i 에 가장 합리적인 예측 값은 무었일까? 가장 합리적인 예측값은 주어진 x_i 에서 반응변수의 평균값인 $E(y_i|x_i)=\beta_0+\beta_1x_i$ 이다. 여기서 실제 관측하여 얻어진 값 y_i 와 예측값 $\beta_0+\beta_1x_i$ 사이에는 오차에 의해서 차이가 발생할 수 있다. 그 차이를 잔차 (residual)라고 하며 x_i 라고 표기한다.

$$r_i = y_i - E(y_i|x_i) = y_i - (\beta_0 + \beta_1 x_i)$$

잔차는 위에 식에서 알 수 있듯이 관측값과 회귀식을 통한 예측값의 차이를 나타낸 것이다. 그러면 자료를 가장 잘 설명할 수 있는 회귀직선을 얻기 위해서는 잔차 r_i 를 가장 작게하는 회귀모형을 세워야 한다. 잔차들을 최소로 하는 방법들 중 하나인 최소제곱법은 잔차들의 제곱합을 최소로 하는 회귀계수 β_0 와 β_1 를 추정하는 방법이다. 잔차들의 제곱합은 다음과 같이 표현된다.

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$$
(1.3)

i 노트

식 1.3 를 잔차제곱합(residusl sum of square)이라고 부른다. 일반적으로 회귀계수의 값이 특정지어져서 실제로 잔차를 계산할 수 있는 경우 잔차제곱합이라고 부른다. 뒤에 분산분석에서는 잔차제곱합을 SSE(sum of square error)라고 부른다.

잔차제곱합을 최소로 하는 회귀계수의 값을 찾는 최적화의 목표로 잔차제곱합이 제시될 때 이를 오차제곱합 (error sum of square)이라고 부른다.

위의 오차제곱합 $S(\beta_0,\beta_1)$ 을 최소화하는 β_0 와 β_1 의 값을 구하는 방법은 오차제곱합이 β_0 와 β_1 의 미분 가능한 2차 함수이고 아래로 볼록한 함수(convex function)임을 이용한다.

```
gridnum <- 60
sizing <- 5
extrascale <- 10
extrascale2 <- 0.7
b0 <- seq(-17.6-sizing*extrascale, -17.6+sizing*extrascale, length=gridnum )
b1 <- seq(4-sizing*extrascale2, 4+sizing*extrascale2, length=gridnum )

SSE <- matrix(0, gridnum, gridnum )
for (i in 1:gridnum ) {
   for (j in 1:gridnum ) {
      r <- cars$dist- b0[i] -b1[j]*cars$speed</pre>
```

```
SSE[i,j] <- (sum(r^2))/1000
}

persp3D(b0, b1, SSE, theta =10, phi = 20, expand = 1)

## Interactive 3d graph
#fig <- plot_ly(z = ~SSE)
#fig <- fig %>% add_surface()
#fig
```

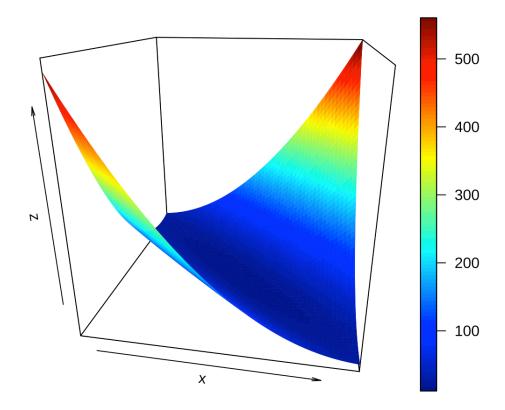


그림 1.1.: 오차제곱합의 함수 형태

위의 그림을 보면 볼록한 모양이 너무 평평하여 오차제곱합이 최소가 되는 β_0 와 β_1 의 위치가 명확하지 않다. 이제 모든 변수들을 표준화하고 표준화된 변수들에 단순회귀모형에 대한 오차제곱합을 β_0 와 β_1 의 함수로서 그림을 그리면 아래와 같다.

$$v_i = \beta_0 + \beta_1 w_i + e_i, \quad i = 1, 2, \dots, n$$
 (1.4)

여기서

$$v_i = \frac{y_i - \bar{y}}{s_y}, \quad w_i = \frac{x_i - \bar{x}}{s_x}$$

```
# 변수들을 표준화!
std_cars <- as.data.frame(scale(cars))</pre>
gridnum <- 60
sizing <- 1
b0 <- seq(0-sizing, 0+sizing, length=gridnum)
b1 <- seq(1-sizing, 1+sizing, length=gridnum )</pre>
SSE <- matrix(0, gridnum, gridnum )</pre>
for (i in 1:gridnum ) {
  for (j in 1:gridnum ){
   r <- std_cars$dist- b0[i] -b1[j]*std_cars$speed
    SSE[i,j] \leftarrow sum(r^2)
  }
}
persp3D(b0, b1, SSE, theta =40, phi = 15, expand = 1)
## Interactive 3d graph
#fig <- plot_ly(z = ~SSE)
#fig <- fig %>% add_surface()
#fig
```

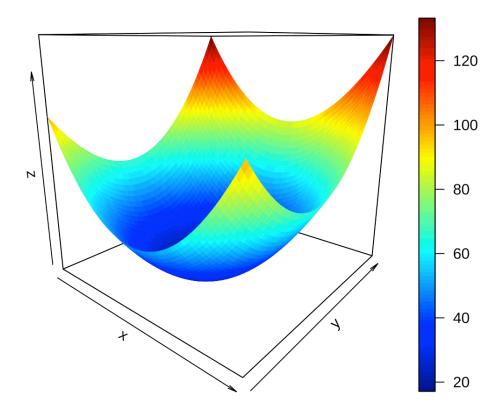


그림 1.2.: 표준화 시 오차제곱합의 함수 형태

위위 같이 변수들을 표준화하면 오차제곱합 함수의 볼록한 정도가 덜 평평하게 변하여 최적값을 더 확실하게 보인다. 기계학습이나 인공지능 모형에서 적합하기 전에 모든 변수를 표준화하는 이유가 위의 그림에서 나타난다.

식 1.3 의 오차제곱합을 각 회귀계수에 대해서 편미분을 하고 0으로 놓으면 아래와 같이 두 방정식이 얻어진다.

$$\begin{split} \frac{\partial S(\beta_0,\beta_1)}{\partial x}\beta_0 &= \sum_{i=1}^n (-2)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \\ \frac{\partial S(\beta_0,\beta_1)}{\partial x}\beta_1 &= \sum_{i=1}^n (-2x_i)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \end{split}$$

위의 연립방정식을 행렬식으로 표시하면 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

위의 방정식을 풀어서 구한 회귀계수의 추정치를 $\hat{eta}_0,\,\hat{eta}_1$ 이라고 하면 다음과 같이 주어진다.

$$\begin{split} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \end{split}$$

최소제곱법에서 얻어진 회귀계수의 추정량 \hat{eta}_0 과 \hat{eta}_1 을 이용한 반응변수 y_i 에 대한 예측값 \hat{y}_i 는 다음과 같이 정의되고

$$\hat{y}_i = \hat{E}(y_i|x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

♦ 표준화 전과 후

두 개의 회귀방정식 식 1.2 과 식 1.4 에서 각각 최소제곱법으로 구한 기울기의 추정치 \hat{eta}_1 이 동일하게 나타 나는 경우는 어떤 경우일까 생각해보자.

잔차 r_i 는 다음과 같이 계산한다.

$$r_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = y_i - \hat{y}_i \tag{1.5}$$

잔차 r_i 는 다음과 같은 성질을 가진다.

$$\sum_{i=1}^n r_i = 0$$

$$\sum_{i=1}^{n} x_i r_i = 0$$

이제 위에서 본 cars 자료를 가지고 선형회귀모형 식 1.2 에 나타난 회귀계수를 추정해보자. 아래는 R 프로그램에 서 함수 1m을 이용한 추정결과이다.

Call:

lm(formula = dist ~ speed, data = cars)

Residuals:

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -17.5791 6.7584 -2.601 0.0123 *

speed 3.9324 0.4155 9.464 1.49e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

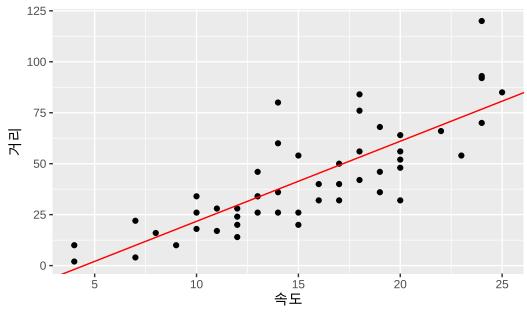
Residual standard error: 15.38 on 48 degrees of freedom Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438

F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

위에서 주어진 선형회귀모형 식 1.2 에 대한 추정 결과를 이용하면 자동차의 속도(x = speed)와 제동거리(y = dist)의 관계는 다음과 같은 회귀식으로 나타낼 수 있다.

$$\hat{E}(y|x) = -17.58 + 3.93x$$

자동차의 속도와 제동거리의 관계



위의 추정식을 이용하면 주어진 자동차의 속도에서 제동거리를 예측할 수 있다. 예를 들어 자동차의 속도가 25 mph 인 경우에는 제동거리의 평균이 80.73 mph 임을 알 수 있다.

$$E(y|x=25) = -17.58 + 3.93(25) = 80.73$$

```
newcars <- data.frame(speed = c(25))
predict(lm_car, newdata=newcars)</pre>
```

1 80.73112

기울기의 추정값 $\hat{\beta}_1=3.93$ 은 자동차의 속도 (x)가 1 mph 증가할 때 평균 제동거리 (E(y|x))가 3.93 ft 증가한 다는 의미이다.

이제 아파트 거애 가격에 대한 단순선형회귀모형을 적합해보자. 이 경우 면적과 가격대신 각각의 로그값을 사용하여 회귀모형을 적합해 보자. 아래는 아파트의 면적과 거래가격에 대한 단순선형 회귀모형을 적합한 결과이다.

```
apart_2019_log <- apart_2019 %>%
  mutate(log_area = log10(area), log_price = log10(price)) %>%
  dplyr::select(log_area, log_price)

head(apart_2019_log,10)
```

```
log_area log_price
```

- 1 1.813514 2.653213
- 2 1.754807 2.440909
- 3 1.960185 2.777427
- 4 1.960185 2.748188
- 5 1.780101 2.628389
- 6 1.780101 2.623249
- 7 1.590173 2.336460
- 8 1.790567 2.685742
- 9 1.856729 2.515874
- 10 1.926548 2.641474

```
lm_apart <- lm( log_price~ log_area, data=apart_2019_log)
summary(lm_apart)</pre>
```

Call:

```
lm(formula = log_price ~ log_area, data = apart_2019_log)
```

Residuals:

```
Min 1Q Median 3Q Max -0.45348 -0.12132 -0.04075 0.07531 0.62358
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.76902 0.05217 14.74 <2e-16 ***

log_area 1.08797 0.02843 38.27 <2e-16 ***

---

Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
```

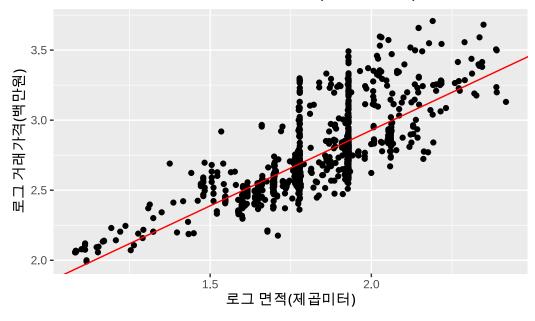
Residual standard error: 0.1894 on 998 degrees of freedom Multiple R-squared: 0.5948, Adjusted R-squared: 0.5944 F-statistic: 1465 on 1 and 998 DF, p-value: < 2.2e-16

위의 결과는 다음과 같이 나타낼 수 있다.

$$\hat{E}(\log 10y|x) = 0.769 + 1.0797 \log 10(x)$$

```
ggplot(apart_2019_log, aes(x=log_area, y=log_price)) + geom_point() + labs(x = "로그 면적(제곱미터 labs(title = "아파트의 면적과 거래가격의 관계(로그스케일)") + geom_abline(intercept = 0.769, slope = 1.0797, color = "red")
```

아파트의 면적과 거래가격의 관계 (로그스케일)



이제 위의 결과를 응용하면 아파트의 면적이 100 제곱미터인 경우의 아파트의 평균 거래가격을 880(백만원)으로 예측할 수 있다.

```
newapart <- data.frame(log_area = c(log10(100)))
pred_y <- 10^predict(lm_apart, newdata=newapart)
pred_y</pre>
```

1

880.9605

1.4. 결정계수

고려한 설명변수와 반응변수에 대하여 제시된 회귀식을 적합한 후 회귀모형이 두 변수의 관계를 얼마나 잘 설명하는지에 대한 기준이 필요하다. 회귀식의 적합에 대한 기준으로서 결정계수(coefficient of determination; R^2)가 있다. 결정계수는 적합의 정도(degree of fitting)를 측정한다. 즉 "설명변수는 반응변수를 얼마나 잘 예측하느냐"에 대한 정도를 수치로 표현한 것이다.

회귀분석에서 설명변수와 반응변수 간에 전혀 관계가 없다면 당연히 반응변수의 값은 설명변수 값의 변동 여하에 전혀 영향을 받지 않아야 한다. 단순회귀모형에서 설명변수 x의 값의 변화를 반응변수 y로 값으로 표현하는것이 바로 기울기 β_1 이다. 이렇게 고려한 설명변수 x가 반응변수 y를 예측하는데 전혀 소용이 없다면 이는 기울기에 대한 회귀계수가 0 $\beta_1=0$ 이라는 것을 의미이다. 이러한 경우에 대하여 다음과 같은 모형을 생각할 수 있다.

$$y_i = \beta_0 + e_i, \quad e_i \sim (0, \sigma^2)$$
 (1.6)

기울기에 대한 회귀계수가 0인 경우에 대한 모형을 4 1.6 과 같이 표현할 수 있으며 평균 모형(mean model)이라고 부른다. 평균 모형은 우리가 생각할 수 있는 모형 중에서 가장 간단한, 하지만 별로 쓸모없는 모형이라고 할 수 있다.

이러한 평균 모형에 대한 최소제곱법을 적용하여 eta_0 의 추정량을 구하면 추정량 \hat{eta}_0 는 $ar{y}$ 가 된다. 그 이유는 위의 모형에 오차제곱합을 구해보면 다음과 같은 형식이 된다

$$S(\beta_0) = \sum_{i=1}^n [y_i - \beta_0]^2$$

여기서 β_0 에 대하여 최고로 하는 지점을 찾아보면 다음과 같은 방정식을 얻을 수 있다.

$$\frac{\partial S(\beta_0)}{\partial \beta_0} = 0 \Rightarrow \sum_{i=1}^n [y_i - \beta_0] = 0$$

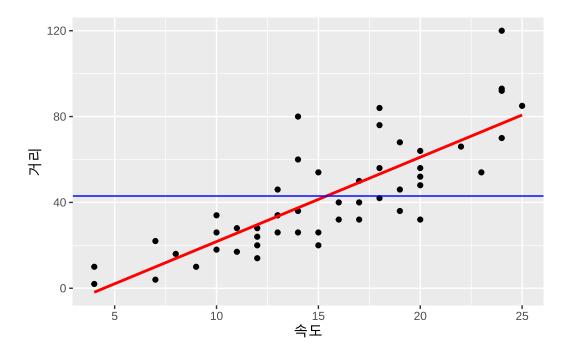
이 방정식을 풀면 $\hat{\beta}_0=\bar{y}$ 가 됨을 알 수 있다. 결국 설명변수가 반응변수에 아무른 영향을 주지 못하게 되면 y의 예측값은 평균 \bar{y} 임을 알 수 있다. 참고로 평균모형 식 1.6 경우 \bar{y} 는 β_0 의 최소제곱추정량이다.

여기서 주목해야할 점은 평균 모형 식 1.6 에서의 잔차 r_{0i} 는 다음과 같이 정의된다.

$$r_{0i}=y_i-\hat{\beta}_0=y_i-\bar{y}$$

주어진 회귀식이 유의한 경우, 즉 회귀식의 기울기가 0이 아닌 경우 $(eta_1 \neq 0)$ 적합된 회귀식에 대한 잔차는 식 1.5 과 같이 나타난다. 만약 회귀식이 유의하다면 식 1.5 으로 구해진 잔차 $r_i = y_i - \hat{eta}_0 - \hat{eta}_1 x_i$ 와 평균 모형에서 구해 지는 잔차 $r_{0i} = y_i - \bar{y}$ 간의 어떤 차이가 있을까?

아래의 그림은 앞의 예제 cars 자료에 대하여 설명변수가 없는 평균 모형(파란 선)과 설명변수가 있는 회귀모형(빨간 선)을 나타낸 그림이다. 잔차는 적합된 직선과 반응 변수 간의 차이를 의미하며 차이의 절대값이 작을 수록 좋은 모형이다.



잔차의 절대값보다 제곱한 양이 다루기가 쉬우므로(why?) 평균 모형과 회귀 모형의 적합도를 비교하는 양으로서 다음과 같은 각각의 모형에서 나온 두개의 잔차제곱합을 생각할 수 있다.

먼저 평균 모형은 예측에 사용할 변수가 없는 경우로서 이때의 잔차는 각 관측값에 대한 예측값이 관측값의 평균이다. 이러한 경우 잔차는 관측값 자체가 가지고 있는 변동으로 생각할 수 있다. 이러한 평균모형에서의 잔차 또는 관측값이 가지고 있는 변동을 총제곱합(Total Sum of Squares; SST)이다

$$\sum_{i=1}^{n} r_{0i}^{2} = \sum_{i=1}^{n} (y_{i} - \bar{y})^{2}$$

$$= \text{Residual Sum of Squares from mean model}$$

$$= \text{Variation of response variables}$$

$$= \text{Total Sum of Squares}$$

$$= SST$$

이제 설명변수가 있는 회귀모형에서 예측치 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$ 를 고려하면 이 경우의 잔차들의 제곱합은 회귀식의 잔차제곱합(Residual Sum of Squares; SSE)이라고 부르며 아래와 같이 정의한다.

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

= Residual Sum of Squares from linear regression model

= Residual Sum of Squares

= SSE

만약 회귀식에서 고려한 설명변수가 반응변수를 예측하는데 매우 적합하다면 회귀모형에서 구한 잔차들의 제곱합이 평균모형에서 구한 잔차들의 제곱합보다 작을 것이다. 이러한 차이를 비교하려면 두 제곱합 SST 와 SSE의 관계를 이해하는 것이 중요하다.

두 제곱합 SST 와 SSE의 관계를 보기 위하여 먼저 두 잔차 r_i^0 와 r_i 의 차이를 비교해 보자

$$r_i^0-r_i=(y_i-\bar{y})-(y_i-\hat{y}_i)=\hat{y}_i-\bar{y}$$

위의 식에서 두 잔차의 차이 $\hat{y}_i - \bar{y}$ 는 예측값과 평균 간 차이로서 그 절대값이 크면 회귀직선이 반응변수를 설명할 수 있는 능력이 크다는 것을 의미한다.

위의 식을 다시 쓰면 다음과 같다.

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

즉, (평균모형의 잔차)=(회귀모형의 잔차))+(회귀모형의 설명부분)으로 분해되는 것으로 이해할 수 있다. 이 분해에서 회귀모형의 잔차가 작을수록 회귀 모형의 예측 능력, 즉 적합도가 커지는 것을 알 수 있다.

이제 총제곱합은 다음과 같이 분해할 수 있다.

$$\begin{split} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 0 \quad \text{(why?)} \end{split}$$

따라서 다음과 같은 제곱합의 분해를 얻게 된다.

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

여기서 모형제곱합(regression sum of square; SSR)를 다음과 같이 정의하면

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2$$

총제곱합은 잔차제곱합 과 모형제곱합으로 분해된다.

$$SST = SSE + SSR \tag{1.7}$$

관측값들이 보여주는 총 변동인 총제곱합(SST)에서 회귀모형으로 설명할 수 있는 변동, 즉 모형제곱합(SSR)이 차지하는 비율을 결정계수(coefficient of determination)라 하며 R^2 으로 표현한다.

$$R^{2} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}$$

위에서 정의된 R^2 는 평균 모형의 잔차제곱합 SST과 회귀모형의 잔차제곱합 SSE의 비율로 정의되는 것으로 해석할 수 있다. 즉,

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\text{Residual SS from regression model}}{\text{Residual SS from mean model}}$$

결정계수의 정의를 보면 회귀모형의 잔차제곱합(SSE)가 평균 모형의 잔차제곱합(SST)에 대하여 **상대적으로** 작아질수록 결정계수가 커진다. 결정계수 R^2 는 언제나 0 이상 1 이하의 값을 갖는다. 회귀모형이 데이터에 아주 잘적합되면 결정계수의 값은 1 에 가깝게 된다.

1.5. 중회귀 모형

일반적으로 회귀모형에서 반응변수의 수는 하나인 경우가 많지만 설명변수의 수는 여러 개인 경우가 많다. 이런 경우 중회귀 모형(multiple linear regression)은 다음과 같이 표현할 수 있고, p-개의 설명변수가 있다고 가정하고 (x_1,x_2,\cdots,x_{n-}) 표본의 크기 n인 자료가 얻어지면 선형회귀식을 행렬로 다음과 같이 표현할 수 있다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{i,p} + e_i = \mathbf{x}_i^t \mathbf{\beta} + e_i$$
 (1.8)

위의 식을 다시 표현하면 다음과 같이 쓸 수 있다.

$$y_i = \pmb{x}_i^t \pmb{\beta} + e_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{i,p} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + e_i$$

이제 n개의 관측치 y_1,y_2,\dots,y_n 으로 이루어진 관측값 벡터 \pmb{y} 를 고려하면 n개의 관측치에 대한 회귀식을 행렬식으로 다음과 같이 표현할 수 있다.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,p} \\ 1 & x_{21} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

위의 식을 간단히 행렬식으로 표시하면 다음과 같다.

$$y = X\beta + e \tag{1.9}$$

위의 행렬식에서 각 벡터와 행렬의 차원은 다음과 같다.

- \mathbf{y} : $n \times 1$
- $X: n \times (p+1)$
- β : $(p+1) \times 1$
- \boldsymbol{e} : $n \times 1$

여기서 회귀분석의 오차항 e_i 은 서로 설명이고 동일한 분산을 갖는다. 즉, 오차항은 다음의 분포를 따른다.

$$\boldsymbol{e} \sim (\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

따라서 관측값 벡터 y의 평균은 다음과 같고

$$E(\mathbf{y}|\mathbf{X}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = \mathbf{X}\boldsymbol{\beta} + E(\mathbf{e}) = \mathbf{X}\boldsymbol{\beta}$$
(1.10)

y의 분산은 아래와 같이 주어진다.

$$V(\boldsymbol{y}|\boldsymbol{X}) = E[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^t] = E(\boldsymbol{e}\boldsymbol{e}^t) = \sigma^2 \boldsymbol{I}_n \tag{1.11}$$

여기서 오차항이 정규분포를 따른다면

$$\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

관측값 벡터 y 또한 정규분포를 따른다

$$\pmb{y} \sim N(\pmb{X}\pmb{\beta}, \sigma^2\pmb{I}_n)$$

1.5.1. 최소제곱추정

이제 중회귀모형 41.9 에서 회귀계수벡터 β 의 추정량을 구하기 위하여 최소제곱법을 적용해보자.

$$\min_{\pmb{\beta}} \sum_{i=1}^{n} (y_i - \pmb{x}_i^t \pmb{\beta})^2 = \min_{\pmb{\beta}} (\pmb{y} - \pmb{X} \pmb{\beta})^t (\pmb{y} - \pmb{X} \pmb{\beta}) \tag{1.12}$$

1.5.1.1. 방법 1

 $\hat{oldsymbol{eta}}$ 는 잔차의 제곱합 식 1.12 을 최소로 하는 최소제곱 추정량이다. 잔차의 제곱합을 $S(oldsymbol{eta})$ 이라고 하면

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^t (\mathbf{y} - \mathbf{X}\beta)$$

$$= \mathbf{y}^t \mathbf{y} - \mathbf{y}^t \mathbf{X}\beta - \beta^t \mathbf{X}^t \mathbf{y} + \beta^t \mathbf{X}^t \mathbf{X}\beta$$

$$= \mathbf{y}^t \mathbf{y} - 2\beta^t \mathbf{X}^t \mathbf{y} + \beta^t \mathbf{X}^t \mathbf{X}\beta$$

여기서 $S(\beta)$ 를 최소로 하는 회귀계수벡터의 값을 구하기 위하여 $S(\beta)$ 를 회귀계수벡터 β 로 미분한후 0 으로 놓고 선형 방정식을 풀어야 한다.

벡터미분을 이용하면

$$\begin{split} \frac{\partial S(\pmb{\beta})}{\partial \pmb{\beta}} &= \frac{\partial}{\partial \pmb{\beta}} (\pmb{y}^t \pmb{y} - 2 \pmb{\beta}^t \pmb{X}^t \pmb{y} + \pmb{\beta}^t \pmb{X}^t \pmb{X} \pmb{\beta}) \\ &= \pmb{0} - 2 \pmb{X}^t \pmb{y} + 2 \pmb{X}^t \pmb{X} \pmb{\beta} \\ &= \pmb{0} \end{split}$$

최소제곱 추정량을 구하기 위한 정규방정식은 다음과 같이 쓸 수 있다.

$$\boldsymbol{X}^{t}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^{t}\boldsymbol{y} \tag{1.13}$$

방정식 식 1.13 를 정규방정식(normal equation)이라고 한다. 만약 $\pmb{X}^t\pmb{X}$ 가 정칙행렬일 경우 최소제곱법에 의한 회귀계수 추정량 $\hat{\pmb{\beta}}$ 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y} \tag{1.14}$$

예측값 벡터 $\hat{\pmb{y}}$ 는 $E(\pmb{y}|\pmb{X})$ 의 추정치로서 다음과 같다.

$$\hat{E}(\boldsymbol{y}|\boldsymbol{X}) = \hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

만약 X^tX 가 정칙행렬이 아닐 경우 최소제곱법에 의한 회귀계수 추정량 $\hat{\boldsymbol{\beta}}$ 은 X^tX 의 일반화 역행렬 $(X^tX)^-$ 를 이용하여 다음과 같이 구한다. 이 경우 일반화 역행렬이 유일하지 않기 때문에 회귀계수 추정량도 유일하지 않다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{X})^- \boldsymbol{X}^t \boldsymbol{y}$$

1.5.1.2. 방법 2

식 1.12 에서 나오는 오차벡터를 정의하고 e=(y-Xeta) 오차벡터를 모수벡터 eta로 미분하면 다음과 같은 결과를 얻는다.

$$\frac{\partial \boldsymbol{e}}{\partial \boldsymbol{\beta}} = \frac{\partial (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{\partial \boldsymbol{X}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \equiv -\frac{\partial \boldsymbol{\beta}^t \boldsymbol{X}^t}{\partial \boldsymbol{\beta}} = -\boldsymbol{X}^t$$

이제 오차제곱합 $S(\pmb{\beta}) = \pmb{e}^t \pmb{e}$ 를 모수벡터로 미분하면 이차형식의 미분공식과 합성함수 미분공식을 차례로 적용하면 된다.

$$\frac{\partial S(\pmb{\beta})}{\partial \pmb{\beta}} = \frac{\partial \pmb{e}^t \pmb{e}}{\partial \pmb{\beta}} = \frac{\partial \pmb{e}}{\partial \pmb{\beta}} \frac{\partial \pmb{e}^t \pmb{e}}{\partial \pmb{e}} = -\pmb{X}^t \left(2\pmb{e} \right) = -2\pmb{X}^t (\pmb{y} - \pmb{X}\pmb{\beta})$$

위의 방정식을 $\mathbf{0}$ 으로 놓으면 최소제곱 추정량 (열)벡터를 구한다.

$$X^t y - X^t X \beta = 0 \quad \rightarrow \quad \hat{\beta} = (X^t X)^{-1} X^t y$$

1.6. 예제-중회귀모형

보기 1.3 (중고차 가격자료). 강의 부교재의 usedcars 자료를 이용하여 중회귀모형을 적합해보자. 자료를 구성하는 변수는 다음과 같다.

• price : 자동차 가격

• yesr :연식

• mileage : 주행거리

• cc : 엔진 크기

• automatic : 자동 변속기 여부

usedcars %>% head(n=10)

automatic	CC	mileage	year	price	
1	1998	133462	78	790	1
1	2000	33000	39	1380	2
0	1800	120000	109	270	3
1	1999	69727	20	1190	4
0	2000	112000	70	590	5
1	1998	39106	58	1120	6

7	815	53	95935	1800	1
8	450	68	120000	1800	0
9	1290	15	20215	1798	1
10	420	96	140000	1800	0

자동차의 가격을 반응변수로 한고 나머지 변수를 설명변수로 설정한 중회귀 모형에 대한 모수의 추정 결과는 다음 과 같다.

```
usedcars_lm <- lm(price ~ year + mileage + cc + automatic, data=usedcars)
summary(usedcars_lm)</pre>
```

Call:

lm(formula = price ~ year + mileage + cc + automatic, data = usedcars)

Residuals:

Min 1Q Median 3Q Max -177.35 -63.91 -0.99 70.34 212.69

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.253e+02 3.998e+02 1.314 0.200823

year -5.800e+00 9.283e-01 -6.247 1.55e-06 ***

mileage -2.263e-03 7.211e-04 -3.138 0.004324 **

cc 3.888e-01 2.022e-01 1.923 0.065958 .

automatic 1.653e+02 3.986e+01 4.147 0.000339 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 101.1 on 25 degrees of freedom

Multiple R-squared: 0.9045, Adjusted R-squared: 0.8892

F-statistic: 59.21 on 4 and 25 DF, p-value: 2.184e-12

1.7. 최소제곱 추정량의 분포

회귀식을 추정하기 위한 회귀계수 추정값인 $\hat{oldsymbol{eta}}$ 의 분포를 알아보기 위해서 우선 선형추정량을 보면 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \equiv \boldsymbol{M}\boldsymbol{y}$$

따라서 최소제곱 추정량은 관측값들의 선형 변환이다. 회귀계수 추정값 $\hat{oldsymbol{eta}}$ 의 기대값은

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{M}\boldsymbol{y}) = E((\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y})$$

$$= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{y})$$

$$= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}$$

$$= \boldsymbol{\beta}$$

따라서 최소제곱 추정량 $\hat{\pmb{\beta}}$ 는 $\pmb{\beta}$ 의 불편추정량이다. 최소제곱 추정량 $\hat{\pmb{\beta}}$ 의 공분산 행렬을 전개해보면

$$\begin{split} Var(\hat{\boldsymbol{\beta}}) &= Var((\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}) \\ &= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\ Var(\boldsymbol{y})\ \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t[\sigma^2\boldsymbol{I}_n]\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1} \end{split}$$

위에서 최소제곱 추정량의 평균과 공분산을 구할 때에는 정규성 가정이 필요하지않다. 만일 y가 정규분포를 따른 다면 y의 선형변환으로부터 얻어진 $\hat{\beta}$ 의 분포는 정규분포이며 다음과 같다.

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^t \boldsymbol{X})^{-1}\right)$$

1.8. 가우스-마코브 정리

정리 1.1 (가우스-마코브 정리). 선형회귀모형 $\pmb{y}=\pmb{X}\pmb{\beta}+\pmb{e}$ 에서 $E(\pmb{e})=0, Var(\pmb{e})=\sigma^2\pmb{I}$ 이 성립하면 최소제곱 추정량

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y}$$

는 β 의 최소분산 선형 불편추정량이다.

위의 정리를 가우스-마코브 정리 (Gauss-Markov Theorem)라고 하며 이는 회귀계수 $\pmb{\beta}$ 의 모든 선형 불편 추정량들 중에 최소제곱 추정량 $\hat{\pmb{\beta}}=(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\pmb{y}$ 이 가장 작은 분산을 가짐을 뜻한다 (Best Linear Unbiased Estimator; BLUE).

가우스-마코브 정리를 정확하게 표현하면 $E(\pmb{L}\pmb{y})=\pmb{\beta}$ 를 만족하는 모든 $n\times n$ 차원의 행렬 \pmb{L} 과 임의의 벡터 \pmb{c} 에 대하여 다음이 성립한다.

$$V(\boldsymbol{c}^t \hat{\boldsymbol{\beta}}) \leq V(\boldsymbol{c}^t \boldsymbol{L} \boldsymbol{y})$$

이제 가우스-마코브 정리를 증명해보자. 관측벡터 y에 대한 임의의 선형 추정량 $\beta^* = Ly$ 를 생각해보면 다시 다음의 형태로 표시할 수 있다.

$$\beta^* = Ly = (M + L - M)y = (M + A)y$$

여기서 $M=(X^tX)^{-1}X^t$ 이고 A=L-M 이다. 임의의 선형 추정량 $oldsymbol{eta}^*$ 가 불편 추정량일 조건을 구해보자

$$E(\boldsymbol{\beta}^*) = E[(\boldsymbol{M} + \boldsymbol{A})\boldsymbol{y}]$$

$$= (\boldsymbol{M} + \boldsymbol{A})E(\boldsymbol{y})$$

$$= (\boldsymbol{M} + \boldsymbol{A})X\boldsymbol{\beta}$$

$$= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}$$

$$= \boldsymbol{\beta} + AX\boldsymbol{\beta}$$

여기서 불편추정량이 되기 위해서는 $E(\pmb{\beta}^*)=\pmb{\beta}$ 조건을 만족 해야되며 따라서 $\pmb{AX}=0$ 이되어야한다 (이 조건은 $\pmb{A}=0$ 를 의미하는 것은 아니다).

이제 최소분산을 가지기 위해서 AX=0을 만족하는 행렬 A중에서 $Var(\pmb{\beta}^*)$ 을 최소로하는 행렬 A를 구해야 한다. $\pmb{\beta}^*$ 의 공분산 행렬은 AX=0이므로

$$\begin{split} V(\pmb{\beta}^*) &= (\pmb{M} + \pmb{A})V(\pmb{y})(\pmb{M} + \pmb{A})^t \\ &= (\pmb{M} + \pmb{A})\sigma^2I_n(\pmb{M} + \pmb{A})^t \\ &= \sigma^2(\pmb{M}\pmb{M}^t + \pmb{A}\pmb{M}^t + \pmb{M}\pmb{A}^t + \pmb{A}\pmb{A}^t) \\ &= \sigma^2[(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\pmb{X}(\pmb{X}^t\pmb{X})^{-1} + \pmb{A}\pmb{X}(\pmb{X}^t\pmb{X})^{-1} + (\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\pmb{A}^t + \pmb{A}\pmb{A}^t] \\ &= \sigma^2[(\pmb{X}^t\pmb{X})^{-1} + \pmb{A}\pmb{A}^t] \\ &= V(\hat{\pmb{\beta}}) + \sigma^2\pmb{A}\pmb{A}^t \end{split}$$

이제 임의의 벡터 c에 대하여

$$egin{aligned} V(oldsymbol{c}^toldsymbol{eta}^*) &= oldsymbol{c}^tV(oldsymbol{eta}^*)oldsymbol{c} \\ &= oldsymbol{c}^tV(\hat{oldsymbol{eta}})oldsymbol{c} + \sigma^2oldsymbol{c}^toldsymbol{A}oldsymbol{A}^toldsymbol{c} \\ &= V(oldsymbol{c}^t\hat{oldsymbol{eta}}) + \sigma^2oldsymbol{c}^toldsymbol{A}oldsymbol{A}^toldsymbol{c} \end{aligned}$$

다음이 성립하므로

$$oldsymbol{c}^t oldsymbol{A} oldsymbol{A}^t oldsymbol{c} = oldsymbol{u}^t oldsymbol{u} = \sum_{i=1}^n u_i^2 \geq 0$$

임의의 벡터 c에 대하여

$$V(\boldsymbol{c}^t \boldsymbol{\beta}^*) \geq V(\boldsymbol{c}^t \hat{\boldsymbol{\beta}})$$

이제 $V(oldsymbol{c}^t oldsymbol{eta}^*)$ 이 $V(oldsymbol{c}^t oldsymbol{\hat{eta}})$ 과 같으려면 다음 조건이 성립해야 하며

$$\boldsymbol{u} = \boldsymbol{c}^t \boldsymbol{A} = \boldsymbol{0}$$

임의의 모든 벡터 c에 대해서 위의 조건 성립해야 하므로 이는 A=0 이 성립해야 한다. 또한 이조건은 AX=0도 만족 시켜준다. 따라서 β 의 최소부산 선형 불편추정량은 최소제곱법으로 구한 추정량이다.

여기서 주의할 점은 가우스-마코브 정리에서 관측값 y에 대한 가정은 평균과 공분산의 가정만 주어졌으며 y의 분포에 대한 가정이 없다. 참고로 만약에 y가 정규분포를 따른다면 최소제곱 추정량은 최소분산 불편추정량이다.

1.9. 최대가능도 추정

관측값 벡터 u 가 다음과 같이 선형모형이며 정규분포를 따른다고 가정하자.

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n) \tag{1.15}$$

선형모형 식 1.15 에 대한 가능도 함수는 다음과 같이 주어진다.

$$\begin{split} L_n(\pmb{\theta}; \pmb{y}) &= L(\pmb{\beta}, \sigma^2 | \pmb{y}) \\ &= \prod_{i=1}^n f(y_i) \\ &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2} (y_i - \pmb{x}_i^t \pmb{\beta})^2\right] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} (\pmb{y} - \pmb{X}\pmb{\beta})^t (\pmb{y} - \pmb{X}\pmb{\beta})\right] \end{split}$$

또한 분산에 대한 모수를 $\tau = \sigma^2$ 과 같이 쓰면 로그 가능도함수는 다음과 같다.

$$\begin{split} \ell_n(\pmb{\theta}; \pmb{y}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{(\pmb{y} - \pmb{X}\pmb{\beta})^t (\pmb{y} - \pmb{X}\pmb{\beta})}{2\sigma^2} \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{(\pmb{y} - \pmb{X}\pmb{\beta})^t (\pmb{y} - \pmb{X}\pmb{\beta})}{2\tau} \end{split}$$

이제 로그가능도함수로부터 구할 수 있는 스코어함수 $s(\pmb{\theta};\pmb{y})$ 와 그에 대한 관측 피셔정보 $J_n(\pmb{\theta};\pmb{y})$ 은 다음과 같이 주어진다.

$$\begin{split} s(\pmb{\theta}; \pmb{y}) &= \frac{\partial}{\partial \pmb{\theta}} \ell_n(\pmb{\theta}; \pmb{y}) \\ &= \begin{bmatrix} \frac{\partial}{\partial \pmb{\beta}} \ell_n(\pmb{\theta}; \pmb{y}) \\ \frac{\partial}{\partial \tau} \ell_n(\pmb{\theta}; \pmb{y}) \end{bmatrix} \\ &= \begin{bmatrix} \pmb{X}^t(\pmb{y} - \pmb{X}\pmb{\beta})/\tau \\ -\frac{n}{2\tau} + \frac{(\pmb{y} - \pmb{X}\pmb{\beta})^t(\pmb{y} - \pmb{X}\pmb{\beta})}{2\tau^2} \end{bmatrix} \end{split}$$

$$\begin{split} J_n(\pmb{\theta}; \pmb{y}) &= -\frac{\partial^2}{\partial \pmb{\theta} \partial \pmb{\theta}} \ell_n(\pmb{\theta}; \pmb{y}) \\ &= - \begin{bmatrix} \frac{\partial^2}{\partial \pmb{\beta} \partial \pmb{\beta}} \ell_n(\pmb{\theta}; \pmb{y}) & \frac{\partial^2}{\partial \pmb{\beta} \partial \tau} \ell_n(\pmb{\theta}; \pmb{y}) \\ \frac{\partial^2}{\partial \tau \partial \pmb{\beta}} \ell_n(\pmb{\theta}; \pmb{y}) & \frac{\partial^2}{\partial \tau \partial \tau} \ell_n(\pmb{\theta}; \pmb{y}) \end{bmatrix} \\ &= \begin{bmatrix} \pmb{X}^t \pmb{X} / \tau & -\pmb{X}^t (\pmb{y} - \pmb{X} \pmb{\beta}) / \tau^2 \\ -(\pmb{y} - \pmb{X} \pmb{\beta})^t \pmb{X} / \tau^2 & -\frac{n}{2\tau^2} + \frac{(\pmb{y} - \pmb{X} \pmb{\beta})^t (\pmb{y} - \pmb{X} \pmb{\beta})}{\tau^3} \end{bmatrix} \end{split}$$

이제 중회귀모형에서 회귀계수 $\pmb{\beta}$ 에 대한 최대가능도 추정량은 스코어함수로 부터 얻어진 방정식 $s(\pmb{\theta};y)=0$ 으로 부터 얻어지며 다음과 같은 형태를 가진다.

$$\hat{\beta} = (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y}$$

$$\hat{\sigma}^2 = \hat{\tau} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^t(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/n = \frac{SSE(\hat{\boldsymbol{\beta}})}{n}$$

여기서 유의할 점은 회귀계수 $m{\beta}$ 의 최대가능도 추정량은 최소제곱법으로 구한 추정량과 동일하다. 따라서 $\hat{m{\beta}}$ 은 최소분산 불편 추정량이다. 하지만 오차항의 분산 σ^2 에 대한 최대가능도 추정량은 불편추정량이 아니다.

$$E(\hat{\sigma}^2) = E\left[(\boldsymbol{y} - \boldsymbol{X} \hat{\boldsymbol{\beta}})^t (\boldsymbol{y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}) / n \right] = E\left[\frac{SSE}{n} \right] \neq \sigma^2$$

참고로 오차항의 분산 σ^2 에 대한 불편추정량은 SSE/(n-p)이다. 오차항의 분산에 대한 불편추정량은 다음 장에서 논의할 것이다.

최대가능도 추정량의 점근적 분포를 이용하면 다음과 같이 말할 수 있다. 오차항이 정규분포인 선형모형인 경우 아래의 분포는 점근분포가 아닌 정확한 분포이다.

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, I_n^{-1}(\boldsymbol{\theta}))$$

여기서

$$I_n(\boldsymbol{\theta}) = E[J(\boldsymbol{\theta}; y)] = \begin{bmatrix} \boldsymbol{X}^t \boldsymbol{X} / \tau & \mathbf{0} \\ \mathbf{0}^t & \frac{n}{2\tau^2} \end{bmatrix}$$

그리고

$$I_n^{-1}(\pmb{\theta}) = \begin{bmatrix} \tau(\pmb{X}^t\pmb{X})^{-1} & \pmb{0} \\ \pmb{0}^t & \frac{2\tau^2}{n} \end{bmatrix} = \begin{bmatrix} \sigma^2(\pmb{X}^t\pmb{X})^{-1} & \pmb{0} \\ \pmb{0}^t & \frac{2\sigma^4}{n} \end{bmatrix}$$

따라서 회귀계수 추정량 $\hat{\beta}$ 의 분포는 평균이 $\pmb{\beta}$ 이고 공분산이 $\sigma^2(\pmb{X}^t-\pmb{X})^{-1}$ 인 정규분포를 따른다. 여기거 주목할 점은 가능도함수에 최대가능도추정량을 대입하면 그 값이 $SSE(\hat{\beta})$ 의 함수로 나타난다.

$$\begin{split} L_n(\hat{\pmb{\theta}}) &= L_n(\hat{\pmb{\beta}}, \hat{\sigma}^2) \\ &= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}(\pmb{y} - \pmb{X}\hat{\pmb{\beta}})^t(\pmb{y} - \pmb{X}\hat{\pmb{\beta}})\right] \\ &= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right] \\ &= \left(2\pi\frac{SSE(\hat{\pmb{\beta}})}{n}\right)^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right] \end{split}$$

또한 가능도함수의 값은 다음과 같다.

$$l_n(\hat{\boldsymbol{\theta}}) = l_n(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = \text{constant} - \frac{n}{2} \log \frac{SSE(\hat{\boldsymbol{\beta}})}{n}$$
 (1.16)

따라서 잔차제곱함 $SSE(\hat{oldsymbol{eta}})$ 작아지면 가능도함수는 커진다.

앞 절에서 언급한 평균 모형 식 1.6 에서 최대가능도 추정을 알아보자. 관측값 벡터는 다음과 같은 분포를 따른다.

$$\boldsymbol{y} \sim N(\beta_0 \mathbf{1}, \sigma^2 \boldsymbol{I}_n) \tag{1.17}$$

선형모형 식 1.15 에 대한 로그 가능도 함수는 다음과 같이 주어진다. 분산에 대한 모수를 $\tau=\sigma^2$ 로 바꾸어 사용하면 모수 벡터는 $\pmb{\theta}=(\beta_0,\tau)^t$ 이다.

$$\ell_n(\boldsymbol{\theta}; \boldsymbol{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{(\boldsymbol{y} - \beta_0 \mathbf{1})^t (\boldsymbol{y} - \beta_0 \mathbf{1})}{2\tau}$$

두 개의 모수 β_0 와 τ 에 대하여 미분하여 가능도 방정식을 구하면 다음과 같다.

$$\begin{split} s(\pmb{\theta}; \pmb{y}) &= \frac{\partial}{\partial \pmb{\theta}} \ell_n(\pmb{\theta}; \pmb{y}) \\ &= \begin{bmatrix} \frac{\partial}{\partial \beta_0} \ell_n(\pmb{\theta}; \pmb{y}) \\ \frac{\partial}{\partial \tau} \ell_n(\pmb{\theta}; \pmb{y}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1}^t(\pmb{y} - \beta_0 \mathbf{1})/\tau \\ -\frac{n}{2\tau} + \frac{(\pmb{y} - \beta_0 \mathbf{1}))^t(\pmb{y} - \beta_0 \mathbf{1})}{2\tau^2} \end{bmatrix} \\ &= \mathbf{0} \end{split}$$

위의 방정식을 풀면 다음과 같은 최대가능도 추정량을 구할 수 있다.

$$\hat{\beta}_0 = \bar{y}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{SST}{n}$$

그리고 가능도함수에 최대가능도추정량을 대입하면 그 값이 다음과 같다.

$$L_n(\hat{\boldsymbol{\theta}}) = L_n(\hat{\beta}_0, \hat{\sigma}^2) = \left(2\pi \frac{SST}{n}\right)^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right]$$
(1.18)

두 개의 모형, 즉 선형회귀모형 식 1.15 과 평균모형 식 1.6 의 가능도 함수의 비, 즉 식 1.16 과 식 1.18 의 비율을 구해보면 결정계수 R^2 외의 관계를 볼 수 있다.

$$\frac{L_n(\hat{\beta_0},\hat{\sigma}^2)}{L_n(\hat{\pmb{\beta}},\hat{\sigma}^2)} = \left(2\pi \frac{SST}{n}\right)^{-\frac{n}{2}} / \left(2\pi \frac{SSE}{n}\right)^{-\frac{n}{2}} \propto \left[\frac{SSE}{SST}\right]^{\frac{n}{2}} = \left[1 - R^2\right]^{\frac{n}{2}}$$

2. 선형회귀에서의 추론

2.1. 제곱합의 분포

앞 장의 중회귀 모형 식 1.9 에서 관측값 벡터 \boldsymbol{y} 가 다변량 정규분포 $N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I})$ 를 따를 때 회귀계수의 추정량 $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$ 은 다음과 같은 분포를 따르는 것을 보였다.

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1})$$

반응변수의 추정값을 구하는 식에서 다음과 같은 모자행렬(hat matrix) $\pmb{H} = \pmb{X}(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t$ 을 정의하자. 여기서 중요한 점은 모자행렬은 대칭인 멱등행렬 ($\pmb{H}\pmb{H} = \pmb{H}$)이며 이는 모자행렬이 사영행렬임을 의미한다.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}$$
(2.1)

2.1.1. 잔차제곱합의 분포

이제 제곱합들의 분포를 알아보기로 하자. 먼저 잔차제곱합 SSE를 이차 형식으로 표시해보자.

$$\begin{split} SSE &= \sum_{i=1}^n (y_i - \hat{y}_i) \\ &= (\boldsymbol{y} - \boldsymbol{X} \hat{\boldsymbol{\beta}})^t (\boldsymbol{y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{y} - \boldsymbol{H} \boldsymbol{y})^t (\boldsymbol{y} - \boldsymbol{H} \boldsymbol{y}) \\ &= \boldsymbol{y}^t (\boldsymbol{I} - \boldsymbol{H})^t (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y} \\ &= \boldsymbol{y}^t (\boldsymbol{I} - \boldsymbol{H}) (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y} \\ &= \boldsymbol{y}^t (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y} \end{split}$$

위의 식에서 I-H는 멱등행렬이고 다음이 성립한다.

$$(I - H)X = X - X(X^tX)^{-1}X^tX = 0$$

따라서

$$\boldsymbol{\mu}^t(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\mu} = \boldsymbol{\beta}^t\boldsymbol{X}^t(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} = 0$$

이므로 비중심 모수는 0이다.

또한

$$\begin{split} r(\pmb{I} - \pmb{H}) &= tr(\pmb{I} - \pmb{H}) \\ &= tr(\pmb{I}_n) - tr\left[\pmb{X}(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\right] \\ &= n - tr\left[(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\pmb{X}\right] \\ &= n - tr(\pmb{I}_p) \\ &= n - p \end{split}$$

이므로 부록의 정리에 의하여 SSE는 다음과 같이 중심 카이제곱 분포를 따른다.

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-p) \tag{2.2}$$

2.1.2. 회귀제곱합의 분포

다음으로 회귀제곱합 SSR의 분포를 유도해보자.

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})$$

$$= (X\hat{\boldsymbol{\beta}} - \bar{y}\mathbf{1})^t (X\hat{\boldsymbol{\beta}} - \bar{y}\mathbf{1})$$

$$= (X\hat{\boldsymbol{\beta}} - \mathbf{1}(\mathbf{1}^t \boldsymbol{y})/n)^t (X\hat{\boldsymbol{\beta}} - \mathbf{1}(\mathbf{1}^t \boldsymbol{y})/n)$$

$$= (H\boldsymbol{y} - \frac{1}{n}\mathbf{1}\mathbf{1}^t \boldsymbol{y})^t (H\boldsymbol{y} - \frac{1}{n}\mathbf{1}\mathbf{1}^t \boldsymbol{y})$$

$$= \boldsymbol{y}^t (H - \frac{1}{n}\boldsymbol{J})^t (H - \frac{1}{n}\boldsymbol{J}) \boldsymbol{y}$$

$$= \boldsymbol{y}^t (H - \frac{1}{n}\boldsymbol{J}) (H - \frac{1}{n}\boldsymbol{J}) \boldsymbol{y}$$

$$= \boldsymbol{y}^t (H - \frac{1}{n}\boldsymbol{J}) \boldsymbol{y}$$

위의 유도식에서 다음 두 가지 성질을 이용하였다. 첫 번째 성질은 모자행렬이 사영행렬이며 모자행렬이 투영하는 공간은 일벡터 $\mathbf{1}$ 을 포함한 공간이다. 이는 계획 행렬 \mathbf{X} 의 첫 번째 열이 절편에 대한 값으로 모두 $\mathbf{1}$ 인 것 때문이다. 따라서

$$egin{aligned} HJ &= H11^t \ &= [H1]1^t \ &= [X(X^tX)^{-1}X^t1]\,1^t \ &= 11^t \ &= J \end{aligned}$$

두 번째로 다음과 같이 JJ = nJ이므로 $\frac{1}{n}J$ 는 멱등행렬이다.

2. 선형회귀에서의 추론

$$egin{aligned} m{J}m{J} &= m{1}m{1}^tm{1}m{1}^t \ &= m{1}[m{1}^tm{1}]m{1}^t \ &= nm{1}m{1}^t \ &= nm{J} \end{aligned}$$

노트

참고로 평균모형 식 1.6 에서 X = 1으므로 이 경우 모자행렬이 다음과 같다.

$$H_0 = \mathbf{1}(\mathbf{1}^t \mathbf{1})^{-1} \mathbf{1}^t = \frac{1}{n} \mathbf{J}$$

다음으로 비중심 모수를 유도하자.

$$egin{aligned} oldsymbol{\mu}^t \left(oldsymbol{H} - rac{1}{n} oldsymbol{J}
ight) oldsymbol{\mu} &= eta^t ig(oldsymbol{X}^t oldsymbol{H} oldsymbol{X} - rac{1}{n} oldsymbol{X}^t oldsymbol{J} oldsymbol{X} ig) oldsymbol{eta} &= eta^t ig(oldsymbol{X}^t oldsymbol{X} - rac{1}{n} oldsymbol{X}^t oldsymbol{J} oldsymbol{X} oldsymbol{eta} ig) \\ &= eta^t oldsymbol{X}^t \left(oldsymbol{I} - rac{1}{n} oldsymbol{J} ig) oldsymbol{X} oldsymbol{eta} &= eta(oldsymbol{eta}) \end{aligned}$$

또한

$$\begin{split} r\left(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}\right) &= tr(\boldsymbol{H}) - tr\left[\frac{1}{n}\boldsymbol{J}\right] \\ &= p - \frac{1}{n}tr(\mathbf{1}\mathbf{1}^t) \\ &= p - \frac{1}{n}tr(\mathbf{1}^t\mathbf{1}) \\ &= p - \frac{1}{n}n \\ &= p - 1 \\ &= p - 1 \end{split}$$

위의 결과를 종합하면 회귀제곱합 SSR은 다음과 같은 분포를 따른다.

$$\frac{SSR}{\sigma^2} \sim \chi^2(p-1,\lambda^2),\tag{2.3}$$

위에서 비중심 모수는 다음과 같다.

$$\lambda^{2} = \frac{1}{\sigma^{2}} \delta(\boldsymbol{\beta}) = \frac{1}{\sigma^{2}} \boldsymbol{\beta}^{t} \boldsymbol{X}^{t} \left(\boldsymbol{I} - \frac{1}{n} \boldsymbol{J} \right) \boldsymbol{X} \boldsymbol{\beta}$$
 (2.4)

2.1.3. 잔차제곱합과 회귀제곱합의 독립

잔차제곱합과 회귀제곱합에서 나타난 이차형식의 두 멱등행렬의 곱은 0이다.

$$(I - H) (H - \frac{1}{n}J) = H - \frac{1}{n}J - HH + \frac{1}{n}HJ$$

= $H - \frac{1}{n}J - H + \frac{1}{n}J$
= 0

따라서 부록의 정리에 의하여 잔차제곱합(SSE)과 회귀제곱합(SSR)은 서로 독립이다.

2.1.4. 총제곱합의 분포

총제곱합 SST의 분포는 위의 결과들을 이용하면 쉽게 구할 수 있다.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \boldsymbol{y}^t \left(\boldsymbol{I} - \frac{1}{n} \boldsymbol{J} \right) \boldsymbol{y}$$

위의 결과를 종합하면 회귀제곱합 SST은 다음과 같은 분포를 따른다.

$$\frac{SST}{\sigma^2} \sim \chi^2(n-1,\lambda^2),\tag{2.5}$$

위에서 비중심 모수 λ^2 은 식 식 2.4 과 같다.

2.2. 모분산의 추정

최소제곱법을 통해서 회귀분석을 실시하였을때 우리는 적합된 회귀선이 얼마나 실제 관측값들을 잘 설명하고 있는 지를 파악하는 것이 모형의 유용성을 판단하는데 중요한 작업이다. 즉, 적합된 회귀선이 관측값을 예측할 때의 변동성을 측정하는 것이 중요하다. 그 변동의 정도를 나타내는 것이 모분산 σ^2 의 추정이다.

식 식 2.2 에 나타난 잔차제곱합의 분포를 이용하면 다음과 같은 결과를 얻는다.

$$E\left[\frac{SSE}{\sigma^2}\right] = n - p$$

위의 방정식에 적률법(Method of Moments)를 적용하면 모분산 σ^2 에 대한 불편추정량을 얻을 수 있다. 평균 잔차 제곱합(mean residual sum of square; S^2 또는 MSE)를 다음과 같이 정의하자.

$$MSE = \frac{SSE}{n-p} = \frac{\sum r_i^2}{n-p} = \frac{\sum (y_i - \hat{y}_i)^2}{n-p} \equiv s^2$$
 (2.6)

 $S^2 = MSE$ 은 모분산의 불편 추정량이다.

2. 선형회귀에서의 추론

$$E(s^2) = E(MSE) = \sigma^2$$

모분산의 추정량이 작을수록 관측값 y의 변동 중 회귀식이 설명할 수 변동이 크다는 것을 나타낸다. 관측값들이 회 귀식으로부터 멀리 떨어져 있으면 MSE 는 커진다.

회귀계수들의 공분산을 추정하는 경우에도 s^2 이 사용된다.

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\boldsymbol{X}^t \boldsymbol{X})^{-1} = s^2 (\boldsymbol{X}^t \boldsymbol{X})^{-1}$$

2.3. 최소제곱 추정량의 성질

최소제곱 추정량의 분포에 대한 성질은 다음과 같다.

- $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1})$
- $\hat{\boldsymbol{\beta}}$ 와 SSE는 독립이다.
- 잔차제곱합(SSE)과 회귀제곱합(SSR)은 서로 독립이다.
- SSE/σ^2 는 자유도가 n-p인 카이제곱 분포를 따른다.
- $(\hat{\pmb{\beta}} \pmb{\beta})^t (\pmb{X}^t \pmb{X}) (\hat{\pmb{\beta}} \pmb{\beta})/\sigma^2$ 는 자유도가 p 인 카이제곱분포를 따른다.

2.4. 모형의 적합도 검정과 분산분석

회귀식을 적합하고 가장 먼저 고려해야할 사항은 적합된 회귀식이 유의한 의미를 가지는지 알아보는 것이다. 회귀식이 가지고 있는 의미는 설명변수의 변화에 따라서 반응변수가 변한다는 것이다. 따라서 회귀 모형이 유의하다는 것은 최소한 하나 이상의 설명변수가 반응변수의 변화를 예측하는데 의미가 있다는 것을 뜻한다. 모든 회귀계수의 값이 0이면 반응변수를 예측하는데 모든 설명변수가 필요가 없다는 것을 의미한다. 이러한 무의미한 모형은 앞장에서 나온 평균모형 식 1.6 이다.

이제 제시된 회귀식이 유의한 지에 대한 검정은 다음과 같은 두 가설 중 하나를 선택하는 것이다.

$$H_0: {\rm mean\ model}\quad vs.\quad H_1:\ {\rm not}\ H_0$$

위의 가설을 바꾸어 쓰면 선형 회귀모형의 유의성 또는 적합도을 검정하는 가설이 된다.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$
 vs. $H_1:$ At least one of β_i is not equal to 0 (2.7)

위의 가설 식 2.7 를 검정하는 방법이 분산분석표를 이용한 F-검정이다.

가설 식 2.7 에서 귀무가설 H_0 가 참인 경우는

2. 선형회귀에서의 추론

$$m{X}m{eta} = [m{1} \ m{x}_1 \ \dots \ m{x}_{p-1}] egin{bmatrix} eta_0 \ 0 \ dots \ 0 \end{bmatrix} = eta_0 m{1}$$

이 성립하여 식 식 2.4 에 나타난 비중심 모수가 0이 된다.

$$\lambda^2 = rac{1}{\sigma^2}oldsymbol{eta}^toldsymbol{X}^t\left(oldsymbol{I} - rac{1}{n}oldsymbol{J}
ight)oldsymbol{X}oldsymbol{eta} = rac{eta_0^2}{\sigma^2}oldsymbol{1}^t\left(oldsymbol{I} - rac{1}{n}oldsymbol{J}
ight)oldsymbol{1} = oldsymbol{0}$$

따라서 귀무가설에서는 회귀제곱합이 자유도가 p-1인 중신p-1인 중신p-1인 공신p-1인 공신p-1인 공기제곱 분포를 따르게 되고 잔차제곱합과 독립 이므로 다음의 통계량 p-1가 자유도가 p-1가고 p-1를 가지는 p-1가고 가는 p-1

$$F_0 = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{MSR}{MSE} \sim F(p-1,n-p) \quad \text{under } H_0 \tag{2.8}$$

따라서 위의 검정 통계량의 p-값이 유의수준보다 크면 적합성 검정에 대한 가설 식 2.7 의 귀무가설을 기각한다. 귀무가설의 기각은 회귀모형의 계수 중 적어도 하나는 0이 아니므로 회귀 모형이 유의하다는 의미이다.

위에서 안급한 F-검정을 위한 통계량들은 다음과 같은 분산분석(Analysis of Variance; ANOVA) 표를 사용하면 쉽게 계산할 수 있다.

표 2.1.: 적합도 검정을 위한 분산분석표

요인	제곱합	자유도	평균제곱합	F-통계량	p-값
회귀	SSR	p-1	MSR	$F_0 = \frac{MSR}{MSE}$	$P(F>F_0)$
오차	SSE	n-p	MSE		
전체	SST	n-1			

3. 모형의 비교

3.1. 직교하는 설명 변수

다음과 같은 2개의 설명변수 (x_1, x_2) 와 반응변수 y 를 가진 자료(데이터프레임)이 있다고 하자.

```
x1 <- c(1, 1, 1, -1, -1, -1, -1)

x2 <- c(1, -1, 1, -1, 1, -1, 1, -1)

y <- c(2, 5, 3, 4, 6, 9, 5, 10)

df <- data.frame(x1, x2, y)

df
```

```
x1 x2 y
1 1 1 2
2 1 -1 5
3 1 1 3
4 1 -1 4
5 -1 1 6
6 -1 -1 9
7 -1 1 5
8 -1 -1 10
```

이제 위의 자료로 선형회귀모형을 적합해 보자.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \tag{3.1}$$

```
fm1 <- lm(y ~ x1 + x2, data=df)
summary(fm1)$coefficients</pre>
```

```
Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.5 0.3162278 17.392527 0.0000115141

x1 -2.0 0.3162278 -6.324555 0.0014565818

x2 -1.5 0.3162278 -4.743416 0.0051344617
```

이제 모형 식 3.1 에서 각각 x_1 과 x_2 를 제거한 축소된 모형을 적합해보자

$$y_i = \beta_0 + \beta_1 x_{i1} + e_i, \qquad y_i = \beta_0 + \beta_2 x_{i2} + e_i$$
 (3.2)

```
fm21 <- lm(y ~ x1 , data=df)
summary(fm21)$coefficients</pre>
```

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.5 0.6770032 8.124038 0.0001867963
x1 -2.0 0.6770032 -2.954196 0.0254739283
```

```
fm22 <- lm(y ~ x2 , data=df)
summary(fm22)$coefficients</pre>
```

```
Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.5 0.8660254 6.350853 0.0007143845

x2 -1.5 0.8660254 -1.732051 0.1339745962
```

두 개의 독립변수가 있는 모형 식 3.1 에서 하나의 독립변수를 제거해도 남아 있는 독립 변수의 회귀계수 추정량은 모형 식 3.1 과 같은 것을 알 수 있다. 이렇게 여러 개의 독립변수가 있는 모형에서 하나의 변수를 제거해도 다른 복립변수의 추정에 영향을 미치지 않는 경우는 어떤 경우일까?

이제 모형 식 3.1 의 설계행렬(X, design matrix)를 구해서 X^tX 를 구해보자.

```
X <- model.matrix(fm1)
X</pre>
```

```
(Intercept) x1 x2
1
           1 1 1
2
           1 1 -1
3
           1 1 1
           1 1 -1
4
5
           1 -1 1
6
           1 -1 -1
7
           1 -1 1
           1 -1 -1
attr(,"assign")
[1] 0 1 2
```

t(X) %*% X

(Intercept) x1 x2 (Intercept) 8 0 0 x1 x2 0 0 8

모형 43.1 의 설계행렬 X의 각 열들은 서로 직교하는것을 알 수 있다.

만약 여러 개의 독립변수를 가진 선형모형에서 모든 설명 변수들의 열들이 모두 서로 직교한다면(절편에 대한 열도 포함해서) 회귀계수의 추정값은 독립변수가 줄어든 축소된 모형에서도 원래의 모형과 같은 것을 알 수 있다.

선형 모형에서 설계행렬 \pmb{X} 의 각 열벡터를 각각 $\pmb{x}_1,\dots,\pmb{x}_p$ 라고 하자. 만약 모든 열들이 서로 직교한다면 (즉 $\pmb{x}_i^t\pmb{x}_i=0$ for $i\neq j$) 선형회귀 모형에서 회귀계수의 추정치는 설명 변수의 유무에 관계없이 일정하게 나타난다.

이러한 상황을 모형식으로 다시 써보자. 만약 다음이 성립하면

$$m{X}^tm{X} = egin{bmatrix} m{x}_1^t \ m{x}_2^t \ dots \ m{x}_p^t \end{bmatrix} egin{bmatrix} m{x}_1 & m{x}_2 & \dots & m{x}_p \end{bmatrix} = egin{bmatrix} m{x}_1^t m{x}_1 & 0 & 0 & \cdots & 0 \ 0 & m{x}_2^t m{x}_2 & 0 & \cdots & 0 \ 0 & 0 & m{x}_3^t m{x}_3 & \cdots & 0 \ dots & dots & dots & dots & dots & dots & dots \ 0 & 0 & \cdots & 0 & m{x}_p^t m{x}_p \end{bmatrix}$$

회귀계수의 추정량은 다음과 같이 나타난다.

$$\begin{pmatrix} \boldsymbol{X}^t \boldsymbol{X} \end{pmatrix}^{-1} \boldsymbol{X}^t \boldsymbol{y} = \begin{bmatrix} \frac{1}{\boldsymbol{x}_1^t \boldsymbol{x}_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\boldsymbol{x}_2^t \boldsymbol{x}_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\boldsymbol{x}_3^t \boldsymbol{x}_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\boldsymbol{x}_1^t \boldsymbol{x}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1^t \boldsymbol{y} \\ \boldsymbol{x}_2^t \boldsymbol{y} \\ \boldsymbol{x}_3^t \boldsymbol{y} \\ \vdots \\ \boldsymbol{x}_p^t \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \frac{\boldsymbol{x}_1^t \boldsymbol{y}}{\boldsymbol{x}_1^t \boldsymbol{x}_1} \\ \frac{\boldsymbol{x}_2^t \boldsymbol{y}}{\boldsymbol{x}_2^t \boldsymbol{x}_2} \\ \vdots \\ \frac{\boldsymbol{x}_p^t \boldsymbol{y}}{\boldsymbol{x}_p^t \boldsymbol{x}_p^t} \end{bmatrix}$$

위의 결과를 조금 더 일반화해보자. 만약 계획행렬 \pmb{X} 를 다음과 같은 p개의 부분 계획행렬 $\pmb{X}_1, \pmb{X}_2, \dots, \pmb{X}_p$ 로 나누고

$$oldsymbol{y} = oldsymbol{X}oldsymbol{eta} + oldsymbol{e} = \sum_{k=1}^p oldsymbol{X}_koldsymbol{eta}_k + oldsymbol{e}$$

부분 계획행렬들이 다음과 같은 성질을 가지고 있다고 하자.

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ \boldsymbol{X}_2 \ \dots \ \boldsymbol{X}_p] \quad \text{and} \quad \boldsymbol{X}_i^t \boldsymbol{X}_j = \boldsymbol{0}, i \neq j$$
 (3.3)

이러한 조건 하에서는 회귀계수의 추정량이 다음과 같이 나타난다.

$$\begin{split} \hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^t \boldsymbol{X} \right)^{-1} \boldsymbol{X}^t \boldsymbol{y} \\ &= \begin{bmatrix} (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} & 0 & 0 & \cdots & 0 \\ 0 & (\boldsymbol{X}_2^t \boldsymbol{X}_2)^{-1} & 0 & \cdots & 0 \\ 0 & 0 & (\boldsymbol{X}_3^t \boldsymbol{X}_3)^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & (\boldsymbol{X}_p^t \boldsymbol{X}_p)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1^t \boldsymbol{y} \\ \boldsymbol{X}_2^t \boldsymbol{y} \\ \boldsymbol{X}_3^t \boldsymbol{y} \\ \vdots \\ \boldsymbol{X}_p^t \boldsymbol{y} \end{bmatrix} \\ &= \begin{bmatrix} (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^t \boldsymbol{y} \\ (\boldsymbol{X}_2^t \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^t \boldsymbol{y} \\ \vdots \\ (\boldsymbol{X}_p^t \boldsymbol{X}_p)^{-1} \boldsymbol{X}_p^t \boldsymbol{y} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\beta}}_3 \\ \vdots \\ \hat{\boldsymbol{\beta}}_p \end{bmatrix} \end{split}$$

위의 결과는 전체 모형에서의 추정량 $\hat{\pmb{\beta}}$ 의 j 번째 부분 $\hat{\pmb{\beta}}_j$ 이 $(\pmb{X}_j^t\pmb{X}_j)^{-1}\pmb{X}_j^t\pmb{y}$ 로 구성되며 다른 \pmb{X}_i 와는 관계가 없다.

이러한 결과를 이용하면 설명변수들이 서로 직교하는 조건 식 3.3를 만족하면 축소모형

$$oldsymbol{y} = oldsymbol{X}_j oldsymbol{eta}_j + oldsymbol{e}$$

에서 계수 추정치 $\hat{\pmb{\beta}}_j = (\pmb{X}_j^t \pmb{X}_j)^{-1} \pmb{X}_j^t \pmb{y}$ 는 모든 설명 변수를 고려한 완전 모형에서의 추정치와 같은 것을 알 수 있다. 설명변수들이 서로 직교하는 조건 식 3.3 을 만족하면 하나의 축소모형에 대한 추정량는 직교하는 다른 설명변수들의 영향을 받지 않는다.

더 나아가서 직교하는 설명변수들이 회귀제곱합에 미치는 영향은 각 축소모형들의 기여도를 단순하게 더한 결과와 같다.

$$\begin{split} SSR &= \boldsymbol{y}^t \left(\boldsymbol{H} - \frac{1}{n} \boldsymbol{J} \right) \boldsymbol{y} \\ &= \boldsymbol{y}^t \boldsymbol{H} \boldsymbol{y} - \frac{1}{n} \boldsymbol{y}^t \boldsymbol{J} \boldsymbol{y} \\ &= \boldsymbol{y}^t \boldsymbol{X} (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y} - n(\bar{y})^2 \\ &= \boldsymbol{y}^t \boldsymbol{X} (\boldsymbol{X}^t \boldsymbol{X})^{-1} (\boldsymbol{X}^t \boldsymbol{X}) (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y} - n(\bar{y})^2 \\ &= \hat{\boldsymbol{\beta}}^t (\boldsymbol{X}^t \boldsymbol{X}) \hat{\boldsymbol{\beta}} - n(\bar{y})^2 \\ &= \sum_{k=1}^p \hat{\boldsymbol{\beta}}_k^t (\boldsymbol{X}_k^t \boldsymbol{X}_k) \hat{\boldsymbol{\beta}}_k - n(\bar{y})^2 \end{split}$$

또한 회귀계수 추정량의 분산을 보면 부분 회귀계수 추정량 $\hat{\pmb{\beta}}_j$ 들은 서로 독립이며 축소 모형에서의 분산과 동일함을 알 수 있다.

$$\begin{split} Cov(\hat{\pmb{\beta}}) &= \sigma^2(\pmb{X}^t\pmb{X}_1)^{-1} & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2(\pmb{X}_2^t\pmb{X}_2)^{-1} & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2(\pmb{X}_3^t\pmb{X}_3)^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2(\pmb{X}_p^t\pmb{X}_p)^{-1} \end{split} \\ &= \begin{bmatrix} Cov(\hat{\pmb{\beta}}_1) & 0 & 0 & \cdots & 0 \\ 0 & Cov(\hat{\pmb{\beta}}_2) & 0 & \cdots & 0 \\ 0 & 0 & Cov(\hat{\pmb{\beta}}_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & Cov(\hat{\pmb{\beta}}_p) \end{bmatrix} \end{split}$$

3.2. 설명변수의 추가

먼저 계획행렬 X_1 을 고려한 선형모형을 고려하자.

$$\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_{1*} + \boldsymbol{e} \tag{3.4}$$

이 경우 회귀계수의 최소제곱추정량은 $\hat{\boldsymbol{\beta}}_{1*} = (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^t \boldsymbol{y}$ 이다.

이제 위의 모형 식 3.4 에 설명변수를 추가한 모형을 생각해 보자. 추가된 설명변수로 이루어진 계획행렬을 \pmb{X}_2 라고 하면 다음과 같이 쓸수 있다.

$$y = X_1 \beta_1 + X_2 \beta_2 + e$$

$$= [X_1 X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + e$$

$$= X\beta + e$$
(3.5)

위의 식에서

$$m{X} = [m{X}_1 \ m{X}_2] \quad ext{and} \quad m{eta} = egin{bmatrix} m{eta}_1 \\ m{eta}_2 \end{bmatrix}$$

설명변수를 추가한 확대 모형 식 3.5 에서 회귀계수의 최소제곱추정량은 $\hat{\pmb{\beta}}=(\pmb{X}^t\pmb{X})^{-1}\pmb{X}^t\pmb{y}$ 이다.

여기서 주의할 점은 식 3.4의 회귀 계수 $\pmb{\beta}_{1*}$ 의 추정량과 식 3.5의 회귀 계수 $\pmb{\beta}_1$ 의 추정량은 일반적으로 같지 않다.

확대모형 식 3.5 의 회귀계수 추정량을 구하려면 최소제곱법을 다시 확장모형에 적용해야 하지만 원래의 모형 식 3.4 에서 구해진 추정량 $\hat{\pmb{\beta}}_{1*}=(\pmb{X}_1^t\pmb{X}_1)^{-1}\pmb{X}_1^t\pmb{y}$ 을 이용하여 유도할 수 있다.

이제 원래의 모형 식 3.4 에서 모자행렬을

$$\boldsymbol{H}_1 = \boldsymbol{X}_1 (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^t$$

라고 하고 확대 모형 식 3.5 에 대하여 다음과 같이 모형을 다시 표현해보자.

$$\begin{split} & \boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \\ & = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + (\boldsymbol{H}_1 + \boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \\ & = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{H}_1 \boldsymbol{X}_2 \boldsymbol{\beta}_2 + (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \\ & = [\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{H}_1 \boldsymbol{X}_2 \boldsymbol{\beta}_2] + (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \\ & = \boldsymbol{X}_1 [\boldsymbol{\beta}_1 + (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^t \boldsymbol{X}_2 \boldsymbol{\beta}_2] + \tilde{\boldsymbol{X}}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \\ & = \boldsymbol{X}_1 [\boldsymbol{\beta}_1 + (\boldsymbol{X}_2^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^t \boldsymbol{X}_2 \boldsymbol{\beta}_2] + \tilde{\boldsymbol{X}}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \end{split}$$

이제 다음과 같은 변화된 모형을 고려하자.

$$\boldsymbol{y} = \boldsymbol{X}_1 \tilde{\boldsymbol{\beta}}_1 + \tilde{\boldsymbol{X}}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} \tag{3.6}$$

위의 식에서 다음과 같이 새로운 계수벡터 $ilde{oldsymbol{eta}}_1$ 와 변환된 계획행렬 $ilde{oldsymbol{X}}_2$ 를 정의하였다.

$$\tilde{\boldsymbol{\beta}}_{1} = \boldsymbol{\beta}_{1} + (\boldsymbol{X}_{1}^{t}\boldsymbol{X}_{1})^{-1}\boldsymbol{X}_{1}^{t}\boldsymbol{X}_{2}\boldsymbol{\beta}_{2}, \quad \tilde{\boldsymbol{X}}_{2} = (\boldsymbol{I} - \boldsymbol{H}_{1})\boldsymbol{X}_{2}$$
 (3.7)

이제 변환된 모형 식 3.6 에서 두 계획행렬 \pmb{X}_1 과 $\tilde{\pmb{X}}_2$ 가 서로 직교하는 것을 알 수 있다.

$$X_1^t \tilde{X}_2 = X_1^t (I - H_1) X_2 = X_1^t (I - X_1 (X_1^t X_1)^{-1} X_1^t) X_2 = 0$$

이제 확대된 모형 식 3.6 의 두 계획행렬 \pmb{X}_1 과 $\tilde{\pmb{X}}_2$ 가 서로 직교하므로 앞에서 나온 직교하는 계획행렬에 대한 회귀계수에 대한 결과를 이용하면 다음과 같이 회귀계수 추정량을 얻을 수 있다.

이제 모형 식 3.6 의 회귀계수 $ilde{oldsymbol{eta}}_1$ 과 $oldsymbol{eta}_2$ 의 추정량을 구해보면 다음과 같다.

$$\hat{\tilde{\beta}}_{1} = (\boldsymbol{X}_{1}^{t} \boldsymbol{X}_{1})^{-1} \boldsymbol{X}_{1}^{t} \boldsymbol{y}, \quad \hat{\boldsymbol{\beta}}_{2} = (\tilde{\boldsymbol{X}}_{2}^{t} \tilde{\boldsymbol{X}}_{2})^{-1} \tilde{\boldsymbol{X}}_{2}^{t} \boldsymbol{y}$$
(3.8)

먼저 식 3.6 에서 $\pmb{\beta}_2$ 에 대한 추정량 $\hat{\pmb{\beta}}_2$ 는 반응변수 \pmb{y} 를 변환된 계획 행렬 $\tilde{\pmb{X}}_2=(\pmb{I}-\pmb{H}_1)\pmb{X}_2$ 로 적합할 때의 회귀계수이다.

$$y = [(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]\boldsymbol{\beta}_2 + \boldsymbol{e}$$

식 3.8 에 주어진 추정량 $\hat{\pmb{\beta}}_2$ 를 다시 다음과 같이 유도할 수 있다.

$$\begin{split} \hat{\pmb{\beta}}_2 &= (\tilde{\pmb{X}}_2^t \tilde{\pmb{X}}_2)^{-1} \tilde{\pmb{X}}_2^t \pmb{y} \\ &= [\pmb{X}_2^t (\pmb{I} - \pmb{H}_1) (\pmb{I} - \pmb{H}_1) \pmb{X}_2]^{-1} \pmb{X}_2^t (\pmb{I} - \pmb{H}_1) \pmb{y} \\ &= [\pmb{X}_2^t (\pmb{I} - \pmb{H}_1) (\pmb{I} - \pmb{H}_1) \pmb{X}_2]^{-1} \pmb{X}_2^t (\pmb{I} - \pmb{H}_1) (\pmb{I} - \pmb{H}_1) \pmb{y} \\ &= (\tilde{\pmb{X}}_2^t \tilde{\pmb{X}}_2)^{-1} \tilde{\pmb{X}}_2^t [(\pmb{I} - \pmb{H}_1) \pmb{y}] \\ &= (\tilde{\pmb{X}}_2^t \tilde{\pmb{X}}_2)^{-1} \tilde{\pmb{X}}_2^t \tilde{\pmb{y}} \end{split}$$

위의 유도를 보면 반응변수 m y 에서 $m X_1$ 로 적합한 후에 구한 잔차벡터 $m y=(m I-m H_1)m y$ 를 새로운 반응변수로 고려한 후, 추가된 변수에 대한 계획행렬 $m X_2$ 에서 먼저 고려한 변수의 계획행렬 $m X_1$ 의 효과를 제거한 $(m I-m H_1)m X_2$ 로 적합한 경우의 회귀계수로 나타난다.

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}_2 \boldsymbol{\beta}_2 + \mathbf{e} \quad \rightarrow \quad (\mathbf{I} - \mathbf{H}_1) y = [(\mathbf{I} - \mathbf{H}_1) \mathbf{X}_2] \boldsymbol{\beta}_2 + \mathbf{e}$$

또한 식 3.6 의 회귀계수 $\tilde{\pmb{\beta}_1}$ 의 추정량은 직교성에 의하여 $(\pmb{X}_1^t\pmb{X}_1)^{-1}\pmb{X}_1^t\pmb{y}$ 으로 주어지며 이는 모형 식 3.4 에서 구한 회귀계수의 추정량과 같다.

$$\hat{\hat{m{eta}}_1} = \hat{m{eta}}_{1*} = (m{X}_1^t m{X}_1)^{-1} m{X}_1^t m{y}$$

이제 식 3.7 의 관계를 이용하면 모형 식 3.5 에서 나타난 회귀계수 $oldsymbol{eta}_1$ 의 추정량을 다음과 같이 표현할 수 있다.

$$\hat{\boldsymbol{\beta}}_{1} = \hat{\tilde{\boldsymbol{\beta}}}_{1} - (\boldsymbol{X}_{1}^{t} \boldsymbol{X}_{1})^{-1} \boldsymbol{X}_{1}^{t} \boldsymbol{X}_{2} \hat{\boldsymbol{\beta}}_{2}
= \hat{\boldsymbol{\beta}}_{1*} - (\boldsymbol{X}_{1}^{t} \boldsymbol{X}_{1})^{-1} \boldsymbol{X}_{1}^{t} \boldsymbol{X}_{2} \hat{\boldsymbol{\beta}}_{2}
= (\boldsymbol{X}_{1}^{t} \boldsymbol{X}_{1})^{-1} \boldsymbol{X}_{1}^{t} (\boldsymbol{y} - \boldsymbol{X}_{2} \hat{\boldsymbol{\beta}}_{2})$$
(3.9)

식 3.9 에 주어진 회귀계수 $m{eta}_1$ 의 추정량은 새로운 변수를 추기하기 전의 모형에서 구한 추정량 $\hat{m{eta}}_{1*}$ 을 새로운 변수를 추가한 후의 모형에서 추가된 변수에 대한 회귀계수의 추정량 $\hat{m{eta}}_2$ 으로 보정힌 형태이다.

이제 간단한 예제를 통하여 위에서 유도한 공식을 적용해 보자.

먼저 3 개의 설명변수 x_1, x_2, x_3 를 가진 10 개의 자료를 임의로 만들어 보자.

```
set.seed(23123)
x1 <- c(1,2,3,4,5,6,7,8,9,9)
x2 <- c(1,4,2,5,3,2,4,3,1,2)
x3 <- c(6,3,2,3,1,4,5,3,2,1)
y <- 2 + 3*x1 + 4*x2 + 5*x3 + rnorm(10)

df <- data.frame(x1,x2,x3,y)
df</pre>
```

3. 모형의 비교

```
x1 x2 x3
  1 1 6 38.99584
1
2
   2 4 3 38.17609
   3 2 2 27.13590
3
4
   4 5 3 49.86576
5
   5 3 1 32.88329
   6 2 4 48.43808
6
7
  7 4 5 63.30271
8 8 3 3 53.54604
   9 1 2 41.86736
9
10 9 2 1 43.30445
```

먼저 2개의 독립변수 x_1 과 x_2 가 있는 모형을 적합해 보자.

$$y = \beta_{0*} + \beta_{1*}x_1 + \beta_{2*}x_2 + e \tag{3.10}$$

lm1s <- lm(y ~ x1 + x2, data=df)
lm1s\$coefficients</pre>

(Intercept) x1 x2 22.964557 1.948430 3.802026

또한 모형 식 3.10 에서 사용한 계획행렬 \pmb{X}_1 을 구해보자

X1 <- model.matrix(lm1s)
X1</pre>

[1] 0 1 2

다음으로 모든 독립변수가 있는 모형을 적합해 보자.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \tag{3.11}$$

lmAll <- lm(y ~ x1 + x2 + x3, data=df)
lmAll\$coefficients</pre>

(Intercept) x1 x2 x3 -0.05735613 3.15491311 4.16172306 5.17857500

또한 모형 식 3.11 에서 모든 독립변수를 사용한 경우 계획행렬 \pmb{X} 와 추가된 변수 x_3 에 대한 열만 가지는 계획행렬 \pmb{X}_2 를 구해보자

X <- model.matrix(lmAll)
X</pre>

X2 <- X[,4]

[1] 0 1 2 3

X2

1 2 3 4 5 6 7 8 9 10 6 3 2 3 1 4 5 3 2 1

이제 식 3.7 에 주어진 $\tilde{\pmb{X}}_2$ 를 계산하고 이를 이용하여 $\hat{\pmb{eta}}_2$ 를 구해보자.

```
H1 <- X1 %*% solve(t(X1) %*% X1) %*% t(X1)

X2t <- (diag(10) - H1) %*% X2

X2t
```

[,1]

- 1 1.8568267
- 2 -0.7018216
- 3 -1.6077630
- 4 -0.1664113
- 5 -2.0723527
- 6 1.0911645
- 7 2.4630575
- 8 0.6265747
- 9 -0.2793667
- 10 -1.2099081

beta2 <- solve(t(X2t) %*% X2t) %*% t(X2t) %*% y beta2

[,1]

[1,] 5.178575

위에서 구한 회귀계수 beta2 는 모든 독립 변수가 있는 모형에서의 x_3 에 대한 회귀계수 추정량과 같다.

이제 절편, x_1 , x_2 만 있는 모형에서 구한 회귀계수를 위에서 구한 beta2를 이용하여 보정해 보자. 아래 보정된 회귀계수 추정량은 모든 독립변수를 고려한 모형에서의 회귀계수 추정량과 같다.

beta1 <- lm1s\$coefficients - solve(t(X1) %*% X1) %*% t(X1) %*% X2 %*% beta2 beta1

[,1]

(Intercept) -0.05735613

x1 3.15491311

x2 4.16172306

이제 앞에서 본 예제와 같이 특별하게 1개의 설명변수를 추가하는 경우를 알아보자. 이 경우는 추가된 변수에 대한 계획행렬 $m{X}_2 = m{x}_p$ 는 하나의 벡터이다. 따라서

$$\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 = \boldsymbol{X}_1 \boldsymbol{\beta} + \boldsymbol{x}_p \boldsymbol{\beta}_p + \boldsymbol{e}$$
(3.12)

위의 식 3.8 에서

$$\tilde{\pmb{X}}_2 = (\pmb{I} - \pmb{H}_1) \pmb{x}_p \equiv \tilde{\pmb{x}}_p$$

로 정의하면 회귀식 식 3.12 에서 하나 추가된 설명변수에 대한 회귀계수의 추정량은 다음과 같다.

$$\begin{split} \hat{\beta}_p &= (\tilde{\boldsymbol{X}}_2^t \tilde{\boldsymbol{X}}_2)^{-1} \tilde{\boldsymbol{X}}_2^t \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \tilde{\boldsymbol{x}}_p^t \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \boldsymbol{x}_p^t (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \boldsymbol{x}_p^t (\boldsymbol{I} - \boldsymbol{H}_1) (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \tilde{\boldsymbol{x}}_p^t (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \tilde{\boldsymbol{x}}_p^t (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y} \\ &= [\tilde{\boldsymbol{x}}_p^t \boldsymbol{x}_p]^{-1} \tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{y}} \\ &= \frac{\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{x}}_p} \end{split}$$

위의 식에서 $\tilde{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y}$ 이다.

또한 식 3.12 과 같이 새로운 변수 x_p 가 추가되면 새로운 변수가 추가된 후에는 회귀계수 추정량이 다음과 같아 보정된다.

$$\begin{split} \hat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1^t \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^{\ t} [\boldsymbol{y} - \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2] \\ &= (\boldsymbol{X}_1^{\ t} \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^{\ t} \boldsymbol{y} - (\boldsymbol{X}_1^{\ t} \boldsymbol{X})_1^{-1} \boldsymbol{X}_1^t \boldsymbol{x}_p \frac{\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{x}}_p} \\ &= (\boldsymbol{X}_1^t \boldsymbol{X})_1^{-1} \boldsymbol{X}_1^t \boldsymbol{y} - (\boldsymbol{X}_1^t \boldsymbol{X})_1^{-1} \boldsymbol{X}_1^t [\boldsymbol{x}_p (\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{x}}_p)^{-1} \tilde{\boldsymbol{x}}_p^t] \tilde{\boldsymbol{y}} \\ &= \hat{\boldsymbol{\beta}}_{1*} - (\boldsymbol{X}_1^t \boldsymbol{X})_1^{-1} \boldsymbol{X}_1^t [\boldsymbol{x}_p (\tilde{\boldsymbol{x}}_p^t \tilde{\boldsymbol{x}}_p)^{-1} \tilde{\boldsymbol{x}}_p^t] \tilde{\boldsymbol{y}} \end{split}$$

3.3. 부분 F-검정과 가능도비 검정

앞 절에서 회귀모형에 새로운 독립변수를 1개 이상 추가할 경우 회귀계수 추정량과 제곱합의 변화를 살펴보있다.

실제 자료를 분석하여 회귀 모형식을 만드는 경우 일반적으로 중요한 몇 개의 설명변수부터 모형에 포함시키고 다른 변수들을 추가한다. 반대로 중요한 변수에 대한 사전 정보가 없다면 가능한 모든 변수를 모두 포함시킨 후에 중요하지 않은 변수들을 제거하기도 한다. 이런 두 가지 방법 모두 축차적으로 변수를 추가 또는 제거하는 방법으로고려하는 모형들이 포함 관계를 가진다.

이렇게 포함관계를 가지는 두 모형을 고려해 보자. 먼저 설명변수의 수가 많은 모형을 최대 모형(full model)이라고 부르자.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots \beta_{p-1} x_{i,p} + \beta_p x_{i,p} + \cdots + \beta_{p+q} + e_i$$

위의 식은 모두 절편을 제외하면 모두 p+q 개의 설명변수를 가진 선형 모형이다. 위의 최대모형을 다음과 같은 행렬식으로 써보자

아래에 정의된 최대

$$y = X\beta + e$$
, $e \sim N(0, \sigma^2 I_n)$ Full model (3.13)

이제 축소된 모형으로 최대모형에서 마자막 q개의 설명변수가 모형에 포함되지 않은 경우를 생각하자.

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p} + \beta_p x_{i,p} + e_i$$

축소모형은 다음과 같은 행렬식으로 표시한다.

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{e}_1, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$
 Reduced Model (3.14)

참고로 최대모형 식 3.13 의 계획행렬 $\pmb{X}=[\pmb{X}_1,\pmb{X}_2]$ 의 차원은 $n\times(p+q+1)$ 이고 축소 모형 식 3.14 의 계획행 렬 \pmb{X}_1 의 차원은 $n\times(p+1)$ 이다.

여기서 유의할 점은 최대 모형은 축소모형을 포함한다는 것이다. 만약 최대모형에서 마지막 q개의 설명변수들에 대한 회귀계수들이 모두 0이면, 즉 $\beta_{p+1}=\cdots\beta_{p+q}=0$ 이면 축소모형이 된다.

교재에서는 추가제곱합을 이용한 부분 F-검정(교과서 p.158-161)을 설명한다. 즉, 다음과 같은 가설을 검정하고자한다.

$$H_0: \beta_{p+1} = \beta_{p+2} = \dots = \beta_{p+q} = 0 \text{ (Reduced)} \quad \text{vs.} \quad H_1: \text{ not } H_0 \text{ (Full)}$$
 (3.15)

위와 같은 가설을 검정하기 위한 부분 ${
m F}$ -검정의 검정 통계량 ${
m F}_0$ 는 다음과 같이 주어진다 (교과서 식 4.4).

$$F_0 = \frac{[SSE(R) - SSE(F)]/(df_R - df_F)}{SSE(F)/df_F}$$

만약 H_0 가 참이면 검정 통계량 F_0 는 자유도가 각각 df_R-df_F 와 df_F 를 가지는 F-분포를 따르므로 이를 이용하여 모형에 새로운 변수를 추가하는 검정을 수행하는 부분 F-검정을 실시할 수 있다.

참고로 위의 식들에서 자유도 df_R 과 df_F 는 다음과 같이 주어진다.

$$df_R = n - (p + q + 1), \quad df_E = n - (p + 1)$$

선형모형에 대한 최대 가능도 추정법은 장 1 에서 설명하였다. 위의 최대 모형과 축소모형에 대한 최대가능도 추정법을 설명하기 위하여 다음과 같은 식을 사용할 것이다.

임의의 벡터 \boldsymbol{v} 에 대하여 노름 $\|\boldsymbol{v}\|^2$ 을 다음과 같아 정의한다.

$$\left\| oldsymbol{v}
ight\|^2 = oldsymbol{v}^t oldsymbol{v}$$

최대모형에 대한 계획행렬 X 에 대한 모자행렬을 이용하여 사영행렬 P 와 Q를 다음과 같이 정의한다.

$$P \equiv H(X) = X(X^tX)^{-1}X^t, \quad Q = I - P$$

또한 축소모형의 계획행렬 X_1 에 대한 모자행렬을 이용하여 사영행렬 P_1 와 Q_1 를 다음과 같이 정의한다.

$$P_1 \equiv H_1(X_1) = X_1(X_1^t X_1)^{-1} X_1^t, \quad Q_1 = I - P_1$$

분산에 대한 모수를 $\tau = \sigma^2$ 과 같이 쓰고 편의상 반응 벡터의 평균을 다음과 같이 표시하자

$$\mu = X\beta$$
, $\mu_1 = X_1\beta_1$

이제 최대모형 식 3.13 에 대한 최대 가능도 추정을 생각해 보자.

$$\begin{split} \ell_n(\pmb{\theta}_F; \pmb{y}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} (\pmb{y} - \pmb{X}\pmb{\beta})^t (\pmb{y} - \pmb{X}\pmb{\beta}) \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} \left\| \pmb{y} - \pmb{X}\pmb{\beta} \right\|^2 \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} \left\| \pmb{y} - \pmb{\mu} \right\|^2 \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} \left\| \pmb{y} - \pmb{P}\pmb{y} + \pmb{P}\pmb{y} - \pmb{\mu} \right\|^2 \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} \left[\left\| \pmb{y} - \pmb{P}\pmb{y} \right\|^2 + \left\| \pmb{P}\pmb{y} - \pmb{\mu} \right\|^2 \right] \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tau - \frac{1}{2\tau} \left[\left\| \pmb{Q}\pmb{y} \right\|^2 + \left\| \pmb{P}\pmb{y} - \pmb{\mu} \right\|^2 \right] \end{split}$$

위의 최대 모형에 대한 로그 가능도 함수 $\ell_n(\pmb{\theta};\pmb{y})$ 에서 $\mu=\pmb{X}\pmb{\beta}$ 에 대한 최대가능도 추정량과 모분산 τ 에 대한 추정량은 다음과 같아 주어진다(장 1 참조)

$$\hat{\mu} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}, \quad \hat{\tau}_F = \frac{1}{n} \|\mathbf{Q}\mathbf{y}\|^2 = \frac{1}{n} SSE(F)$$

따라서 최대 가능도 추정량 $\hat{\mu}$ 와 $\hat{\tau}_F$ 를 로그 가능도 함수에 넣으면 다음과 같은 결과가 주어진다.

$$\max_{\boldsymbol{\mu},\boldsymbol{\tau}} \ell_n(\boldsymbol{\theta}_F; \boldsymbol{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \hat{\tau}_F - \frac{n}{2}$$

위의 결과를 가능도 함수로 표시하면

$$\max_{\boldsymbol{u}, \boldsymbol{\tau}} L_n(\boldsymbol{\theta}_F; \boldsymbol{y}) = (2\pi e)^{-n/2} [SSE(F)/n]^{-n/2}$$
(3.16)

축소모형 식 3.14 에 대해서도 같은 방법으로 최대 가능도 출정량을 구하면 다음과 같이 주어지며

$$\hat{\boldsymbol{\mu}}_1 = \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 = \boldsymbol{P}_1 \boldsymbol{y}, \quad \hat{\boldsymbol{\tau}}_R = \frac{1}{n} \left\| \boldsymbol{Q}_1 \boldsymbol{y} \right\|^2 = \frac{1}{n} SSE(R)$$

로그 가능도 함수에 넣어면 다음과 같은 결과가 주어진다.

$$\max_{\pmb{\mu}_1,\pmb{\tau}} \ell_n(\pmb{\theta}_R; \pmb{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \hat{\tau}_R - \frac{n}{2}$$

위의 결과를 가능도 함수로 표시하면

$$\max_{\boldsymbol{u}, \boldsymbol{\tau}} L_n(\boldsymbol{\theta}_R; \boldsymbol{y}) = (2\pi e)^{-n/2} [SSE(R)/n]^{-n/2}$$
(3.17)

최대 가능도 추정에서 만약 최대 모형 F 가 축소모형 R 을 포함하면 가설 식 3.15 에 대한 검정이 가능하다.

위의 가설 검정에서 두 모형의 가능도 함수 식 3.16 와 식 3.17 의 비, 즉 가능도 비(Likelihood Ratio) Λ 는 다음 과 같이 주어진다.

$$\Lambda = \frac{\underset{\boldsymbol{\mu}_{1}, \boldsymbol{\tau}}{\max} L_{n}(\boldsymbol{\theta}_{R}; \boldsymbol{y})}{\underset{\boldsymbol{\mu}_{T}}{\max} L_{n}(\boldsymbol{\theta}_{F}; \boldsymbol{y})} = \left[\frac{SSE(F)}{SSE(R)}\right]^{n/2}$$
(3.18)

위의 가능도 비 Λ 가 작으면 귀무가설 H_0 을 기각한다.

if
$$\Lambda < c'$$
, then reject H_0

식 식 3.18 에 나타난 가능도 비를 다시 표현해 보자

$$\Lambda = \left\lceil \frac{SSE(F)}{SSE(R)} \right\rceil^{n/2} = \left\lceil 1 + \frac{SSE(R) - SSE(F)}{SSE(F)} \right\rceil^{-n/2} \equiv (1 + F^*)^{-n/2}$$

위의 식을 보면 Λ 가 F^* 에 반비례하므로 다음과 같이 F^* 가 크면 H_0 를 기각할 수 있다. 이제 회귀분석에서 주로 쓰이는 부분 F-검정 통계량을 위의 식에서 나타난 F^* 로 표시해보면 다음과 같이 쓸수 있다.

$$\text{if } F = \frac{n - (p + q + 1)}{q} F^* = \frac{[SSE(R) - SSE(F)]/q}{SSE(F)/[n - (p + q + 1)]} > c, \text{ then reject } H_0 \tag{3.19}$$

위의 식 3.19 에 있는 가설검정의 절차는 추가제곱합을 이용한 부분 F-검정(교과서 p.158-161)과 동일한 검정이다. 따라서 부분 F-검정은 가능도비 검정이다.

4. 모형의 진단

4.1. 등분산성 가정의 위반

일반적인 회귀분석모형에서

$$y = X\beta + e$$

오차항이 다음과 같이 서로 독립이고 등분산성을 만족한다면

$$var(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$$

최소제곱법에 의한 회귀계수 추정량 $\hat{oldsymbol{eta}}$ 다음과 같고

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{y}$$

이는 최소분산선형추정량(BLUE)이다. 만약에 오차항에 대한 가정이 만족하지 않는다면 최소제곱법 추정량 $\hat{\pmb{\beta}}$ 의 최적성이 유지되는가에 대한 질문이 생기게 된다.

여기서 오차항의 분산에 대하여 좀더 일반적인 모형을 생각해보자. 가장 일반적인 모형은 다음과 같은 임의의 양정 치행렬(positive definite matrix) $m{V}$ 가 오차항의 공분산 행렬인 경우이다.

$$Var(\boldsymbol{e}) = \boldsymbol{V}$$

가장 일반적인 경우를 고려하기 전에 전형적인 가정을 약간 벗어나면서 실제 문제에서 흔히 접하는 경우를 생각해 보자.

일단 오차항이 서로 독립이지만 분산이 다른 경우이다.

$$Var(\pmb{e}) = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2]$$

4.1.1. 가중 최소제곱법

이러한 경우에 가중 최소제곱법(Weighted Leadt Square Estimator; WLSE)을 사용하면 최소제곱법 추정량의 최적성을 유지할 수 있다.

오차항의 분산 $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ 을 안다고 가정하면 각 관측값 y_i 를 해당 오차의 표분편차 σ_i 로 나누면 등분산성을 다시 얻을 수 있다.

$$Var(y_i) = Var(\boldsymbol{x}_i^t \boldsymbol{\beta} + e_i) = Var(e_i) = \sigma_i^2$$

$$\Rightarrow Var(y_i/\sigma_i) = Var(\boldsymbol{x}_i^t \boldsymbol{\beta}' + e_i/\sigma_i) = Var(e_i/\sigma_i) = 1$$

이 떄 새로운 관측치 $y_i' = y_i/\sigma_i$ 를 사용하여 최소제곱법을 적용하면

$$\min_{\pmb{\beta}'} \sum_{i=1}^n (y_i' - \pmb{x}_i^t \pmb{\beta}')^2 = \min_{\pmb{\beta}} \sum_{i=1}^n \left[\frac{1}{\sigma_i^2}\right] (y_i - \pmb{x}_i^t \pmb{\beta})^2 \equiv \min_{\pmb{\beta}} \sum_{i=1}^n w_i (y_i - \pmb{x}_i^t \pmb{\beta})^2$$

여기서 $w_i=1/\sigma_i^2$ 이다. 이러한 가중최소제곱법은 각 관측치에 대하여 서로 다른 가중치를 적용하여 최소제곱 추정량을 구하며 위의 경우에는 가중치가 반응값의 분산에 반비례한다. 따라서 분산이 큰 오차항을 가진 반응값의 가중치는 분산이 작은 반응값에 비해 상대적으로 작다. 이러한 가중치와 분산의 관계는 변이가 적은 반응값 근방에서 오차를 더욱 줄이려고 하는 직관적인 생각과 일치한다. 가중치를 적용한 최소제곱 추정량은 다음과 같이 나타낼 수있다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{y}$$

여기서

$$m{W} = \mathrm{diag}\left[rac{1}{\sigma_1^2}, rac{1}{\sigma_2^2}, \dots, rac{1}{\sigma_n^2}
ight] = [Var(m{e})]^{-1} = m{V}^{-1}$$

위에서 본 가중최소제곱법을 오차항이 일반적인 공분산 행렬 $m{V}$ 를 가질 때 적용하면 다음과 같이 목적 함수를 나타낼 수 있으며

$$\min_{\boldsymbol{\beta}}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^t\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$$

가중최소제곱추정량은 다음과 같이 나타낼 수 있다.

$$\hat{\boldsymbol{\beta}}_{x} = (\boldsymbol{X}^{t} \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^{t} \boldsymbol{V}^{-1} \boldsymbol{y}$$

$$(4.1)$$

가중최소제곱추정량은 다음과 같은 성질은 만족한다.

- 가중최소제곱추정량은 불편추정량이다: $E(\hat{\pmb{eta}}_{\star}) = \pmb{eta}$
- 가중최소제곱추정량 $\hat{\pmb{\beta}}_*$ 는 분산이 가장 작은 불편선형 추정량이다 (Gauss-Markov Theorem)
- 가중최소제곱추정량 $\hat{\pmb{\beta}}_*$ 의 분산:

$$Var(\hat{\pmb{\beta}}_*) = (\pmb{X}^t\pmb{V}^{-1}\pmb{X})^{-1}$$

- 가중최소제곱추정량 $\hat{\pmb{\beta}}_*$ 은 \pmb{y} 가 평균이 $\pmb{X}\pmb{\beta}$ 이고 분산이 V인 정규분포에서 $\pmb{\beta}$ 에 대한 최대우도추정량이다.
- 일반최소제곱추정량도 불편추정량이다: $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

가중최소제곱법에서 유의할 점은 오차항의 공분산 행렬 $m{V}$ 에 대하여 모르는 경우 이를 추정해야 하며 가중최소제 곱추정량에 공분산 행렬의 추정치

 $\hat{m{V}}$ 을 사용할 경우 위에서 언급한 최적성은 더 이상 성립하지 않는다.

$$\hat{\pmb{\beta}}_* = (\pmb{X}^t \hat{\pmb{V}}^{-1} \pmb{X})^{-1} \pmb{X}^t \hat{\pmb{V}}^{-1} \pmb{y}$$

더 나아가 공분산행렬 V에 대한 추정은 어려운 문제이므로 그 추정 방법과 통계적 성질을 잘 고려하여 사용해야 한다. 정확한 추정을 위해서 또는 자료의 특성을 이용하여 공분산 행렬 V에 대한 모형을 어느 정 \circ 도 단순화하는 것이 바람직하다. 예를 들어 공분산 행렬 V를 다음과 같이 나타낼 수 있다면 유용할 것이다.

$$\mathbf{V} = \sigma^2 \operatorname{diag}[v_1, v_2, \dots, v_n]$$

여기서 (v_1, v_2, \dots, v_n) 은 알려진 값이고 σ^2 는 추정해야하는 모수이다.

오차항의 등분산성에 대한 가정을 검토하기 위한 방법중 가장 유용하고 간단한 방법은 잔차그림을 이용하는 것이다. 잔차 r_i 와 적합된 값 \hat{y}_i 에 대한 잔차그림을 그려서 잔차의 퍼진 정도가 적합된 값 \hat{y}_i 에 따라 변하면 등분산성에 대한 가정을 의심해봐야 한다. 실제 자료에서 반응값의 분산이 독립변수에 비례하여 나타나는 경우가 많다. 단순회 귀모형을 고려하고

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

오차항이 서로 독립이며 그 분산이 독립변수에 비례한다고 가정하자.

$$Var(e_i) = x_i \sigma^2$$

이러한 경우는 독립변수의 값이 양인 경우이며 독립변수의 값이 커지면 반응값의 분산도 커진다. 이러한 경우 독립 변수와 종속변수의 관계, 회귀식, 잔차그림은 다음과 같이 나타난다.

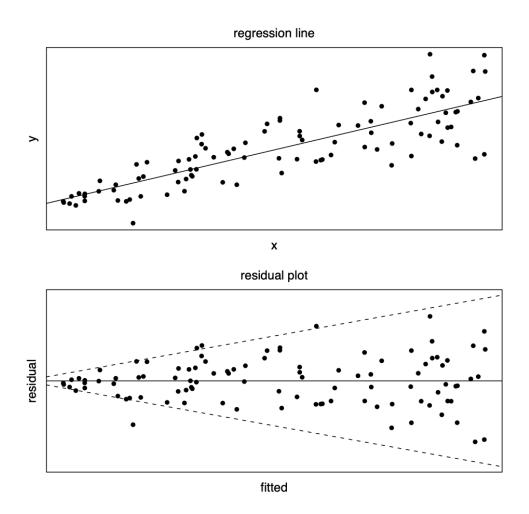


그림 4.1.: 오차항의 등분산성이 위반된 경우

4.2. 변수변환

변수변환(Variable transformation)은 독립변수와 종속변수를 변환함으로서 회귀식의 적합도를 향상시켜 예측 력을 높일 수 있을뿐 아니라 최소제곱법에서의 등분산성 가정에 대한 만족도를 높일 수 있는 유용한 방법이다 (variance stabilization). 이 절에서는 변수변환의 종류와 그 효과를 단순회귀식에서 살펴본다. 중회귀의 경우에는 변수변환의 적용을 복합적으로 고려해야 할 것이다.

4.2.1. 지수모형과 멱함수: 로그변환

회귀식에 대한 모형이 지수함수 모형인 경우, 즉 독립변수와 종속변수가 다음과 같은 경우

$$y = \beta_0 \exp(\beta_1 x)$$

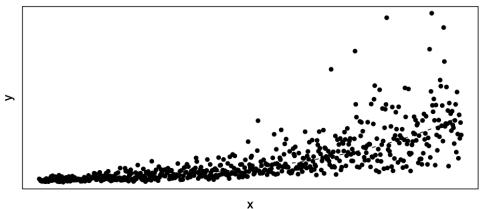
종속변수에 대한 로그 변환(log transformation)을 하면 선형관계에 매우 가깝게 된다 (아래 그림 참조).

$$\log(y_i) = \beta_{i,0} + \beta_1 x_i + e_i$$

여기서 주의할 점은 원래의 지수모형에 오차항의 지수함수가 곱해지는 형태가 되어야 로그 변환후에 등분산성의 가정을 만족하게 된다. 즉 오차항 e_i 를 서로 독립이고 평균이 0, 분산이 σ^2 이라고 하면 다음과 같은 관계가 성립된 다.

$$y_i = \beta_0 \exp(\beta_1 x_i) \exp(e_i) \quad \Rightarrow \quad \log(y_i) = \beta_0' + \beta_1 x_i + e_i$$

Before transfomation: exponential



Log transformation on y

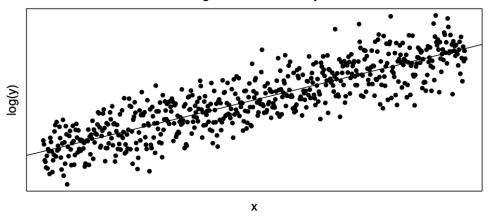


그림 4.2.: 지수모형과 로그변환

회귀식에 대한 모형이 멱함수모형인 경우, 즉 독립변수와 종속변수가 다음과 같은 경우

$$y = \beta_0 x_1^{\beta}$$

독립변수와 종속변수에 대한 로그 변환을 하면 선형관계에 매우 가깝게 된다.

$$\log(y_i) = \beta_{i,0} + \beta_1 \log(x_i) + e_i$$

여기서 주의할 점도 원래의 멱함수모형에 오차항이 곱해지는 형태가 되어야 로그 변환후에 등분산성의 가정을 만족하게 된다. 즉 오차항 e_i 를 서로 독립이고 평균이 0, 분산이 σ^2 이라고 하면 다음과 같은 관계가 성립된다.

$$y_i = \beta_0 x_1^\beta \exp(e_i) \quad \Rightarrow \quad \log(y_i) = \beta_0' + \beta_1 \log(x_i) + e_i$$

회귀식에 대한 모형이 역지수함수 모형인 경우, 즉 독립변수와 종속변수가 다음과 같은 경우

$$y = \beta_0 \exp(\beta_1/x)$$

종속변수에 대한 로그 변환과 독림변수에 대한 역변환을 하면 선형관계에 매우 가깝게 된다.

$$\log(y_i) = \beta,_0 + \beta_1 \left(\frac{1}{x_i}\right) + e_i$$

여기서 주의할 점은 원래의 역지수모형에 오차항의 지수함수가 곱해지는 형태가 되어야 로그 변환후에 등분산성의 가정을 만족하게 된다.

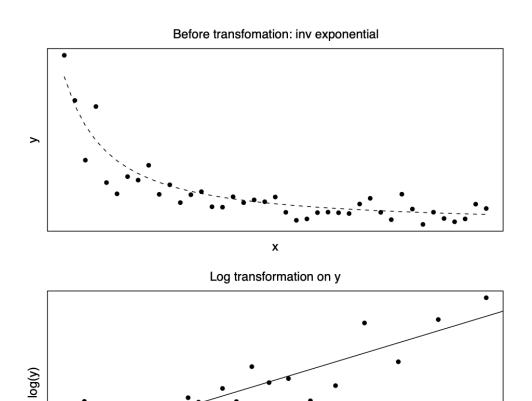


그림 4.3.: 역지수모형과 로그/역변환

1/x

4.2.2. 쌍곡선과 역변환

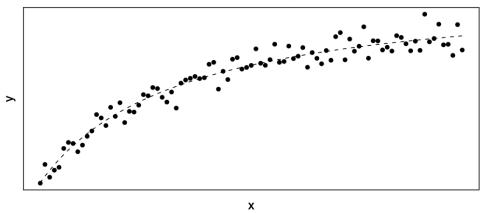
생물학, 경제학등의 문야에서 독립변수와 종속변수의 관계가 쌍곡선(Hyperbola) 형태인 경우가 많다. 독립변수가 증가함에 따라 종속변수의 값이 수렴하는 경우에 이러한 관계가 매우 유용하다.

$$y = \frac{x}{\beta_0 + \beta_1 x}$$

이러한 경우에 독립변수와 종속변수에 모두 역변환(Inverse transformation)을 취하면 선형관계에 매우 가깝게된다 (아래 그림 참조).

$$\frac{1}{y_i} = \beta_0 + \beta_1 \left(\frac{1}{x_i}\right) + e_i$$

Before transfomation: hyperbola



Inverse transformation on x and y

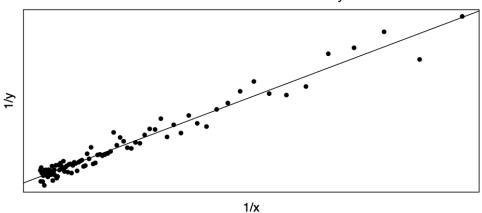
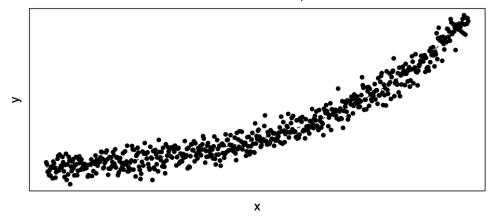


그림 4.4.: 쌍곡선과 역변환

위에서 살펴본 회귀모형들에서 변수변환후에 등분산성에 대한 가정을 만족하려면 대부분 변환전의 함수관계식에 오차항의 지수함수가 곱해지는 형태가 되어야 한다 (multiplicative error model). 이렇게 오차항이 함수관계식에 곱해지는 형태가 아니 다른 형태라면, 예를 들어 오차항이 원래의 함수관계식에 더해지는 형태 (additive error model), 등분산성의 가정이 상당히 위배될 수 있음을 주의해야 한다. 예를 들어 회귀식에 대한 모형이 지수함수 모형인 경우 서로 독립이고 평균이 0, 분산이 σ^2 인 오차항 e_i 를 함수 관계식에 더해졌다면 로그변환된 종속 변수와 독립변수는 선형관계를 보이지만 등분산성 가정은 만족하지 못하게된다.

$$y_i = \beta_0 \exp(\beta_1 x_i) + e_i \quad \Rightarrow \quad \log(y_i) \cong \beta_0' + \beta_1 x_i + e_i^*$$

Before transfomation: exponential



Log transformation on y

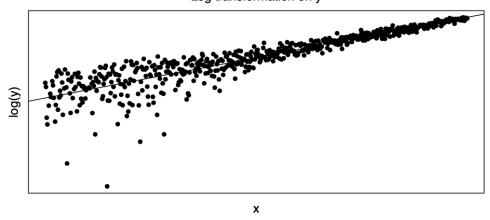


그림 4.5.: Additive error model 과 로그변환

4.2.3. Box-Cox 변화

앞 절에서 보았듯이 종속변수에 로그변환 등을 적용하면 여러 가지 유용한 점이 많다. 위에서 살펴본 종속변수에 변환은 여러 가지 비선형모형을 선형모형에 가깝게 만들어 주며 Multiplicative error model과 같이 반응값이 분산이 독립변수의 크기에 영향을 받는 모형을 등분산성을 가진 형태의 모형으로 버꾸어 준다 (Variance stabilization). 이렇게 종속변수에 대한 여러 가지 변환을 하나의 체계적인 형태로 결합한것을 Box-Cox 변환으로 부르며 다음과 같아 정의한다.

$$y^{(\lambda)} = \begin{cases} \log(y) & \text{if } \lambda = 0\\ \frac{y^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \end{cases}$$
 (4.2)

Box-Cox 변환은 로그변환과 멱변환을 모두 포함하고 있다. 또한 Box-Cox 변환된 $y^{(\lambda)}$ 가 정규분포를 따른다는 가 정 하에 자료가 주어졌을때 λ 에 대한 최대우도추정량을 구할 수 있다.

4.3. 다중공선성

회귀분석에서 독립변수들간의 강한 선형 관계의 경향이 있을 때 이를 다중공선성(multicollinearity)라고 한다. 즉, p개의 독립변수 x_1, x_2, \dots, x_n 의 관계가 다음과 같은 선형관계에 가깝다면 다중공선성이 존재한다고 한다.

$$c_1 x_1 + c_2 x_2 + \dots + c_p x_p \cong 0$$

다중공선성에 의해 발생하는 여러 가지 문제점들을 기술적으로 독립변수들의 강한 선형관계때문에 행렬 X^tX 가 ill-conditioned 행렬이 되어 그 역행렬이 불안정하게 구해지는 결과 때문에 생기게 된다. 여기서 회귀계수의 공분 산 행렬은 다음과 같이 주어짐에 유의하자.

$$Var(\boldsymbol{b}) = \sigma^2 (X^t X)^{-1}$$

예를 들어서 두 개의 독립변수가 있는 회귀 모형을 생각해 보자.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

그림 4.6 에서 (a)의 경우는 두 개의 독립변수 x_1 과 x_2 의 관계가 독립적이어서 (linearly independent) 적합된 회 귀식이(그림에서 2차원 평면) 안정적이다. 반면에 그림 4.6 에서 (b)의 경우는 두 개의 독립변수 x_1 과 x_2 가 완벽한 선형관계가 있기 때문에 (linearly dependent)

$$c_1 x_1 + c_2 x_2 = 0$$

적합된 회귀식이 여러가지 존재한다. 이러한 경우는 매우 드물지만 그림 4.6 에서 (c)의 경우는 두 개의 독립변수 x_1 과 x_2 가 선형관계에 매우 가깝기 때문에 적합된 회귀식이 불안정하다.

회귀분석에서 다중공선성의 정도를 측정할 수 있는 통계량은 다음과 같은 것들이 있다.

4.3.1. 독립변수간의 상관계수

독립변수간의 상관계수를 보아 강한 상관관계를 가지는 변수들이 있다면 다중공선성의 가능성이 크다.

4.3.2. X^tX 의 고유값(Eigenvalues)

행렬 X^tX 의 고유값를 구하여 큰 순서대로 나열했을 때 가장 작은 값이 0에 매우 가까우면 다중공선성의 가능성이 크다. 만약에 독립변수들간의 선형 관계가 있다면 행렬 X^tX 은 최대 계수(full rank) 행렬이 아니므로 하나 이상의 고유값이 0이 되게 된다.

이제 $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ 를 $\boldsymbol{X}^t \boldsymbol{X}$ 의 고유값라고 하자. 이런 가정 하에서 $1/\lambda_i$ 는 $(\boldsymbol{X}^t \boldsymbol{X})^{-1}$ 의 고유값이다. $\boldsymbol{X}^t \boldsymbol{X}$ 의 고유값에 대한 고유벡터를 $\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_p$ 라고 하고 행렬 $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_p]$ 로 정의하자. 이때 다음과 같이 스펙트렇 분해를 이용하여 $\boldsymbol{X}^t \boldsymbol{X}$ 를 나타낼 수 있다.

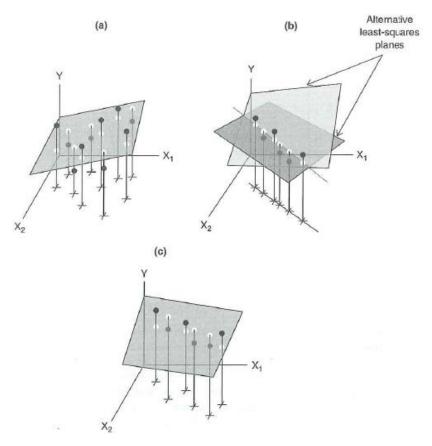


그림 4.6.: 회귀분석에서의 다중공선성의 정도(Fox 2008, p310)

$$\pmb{P}^t(\pmb{X}^t\pmb{X})\pmb{P} = \pmb{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\dots,\lambda_p)$$

또한

$$(\boldsymbol{X}^t\boldsymbol{X})^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^t$$

따라서 가장 작은 고유값 λ_p 가 매우 0에 가까우면 $(\pmb{X}^t\pmb{X})\pmb{p}_p=\lambda_p\pmb{p}_p\approx 0$ 이 성립하고 이는 $\pmb{X}\pmb{p}_p\approx 0$ 을 의미하여 독립변수간에 선형관계가 있다는 것을 암시한다.

일반적으로 최소제곱추정량의 공분산은 다음과 같아 나타내어지고

$$Var(\hat{\pmb{\beta}}) = \sigma^2(\pmb{X}^t\pmb{X})^{-1} = \sigma^2\pmb{P}\pmb{\Lambda}^{-1}\pmb{P}^t = \sigma^2\sum_{i=1}^p\frac{1}{\lambda_i}\pmb{p}_i\pmb{p}_i^t$$

최소제곱추정량의 분산의 합(total variance)은 다음과 같다.

$$\sum_{i=1}^p Var(\hat{\beta}_i) = tr[\sigma^2(\pmb{X}^t\pmb{X})^{-1}] = \sigma^2 tr[(\pmb{X}^t\pmb{X})^{-1}] = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

따라서 고유값의 값이 0에 가까와 지면 추정량의 분산의 합은 매우 크게 된다.

4.3.3. 조건지수 (condition number)

다중공선성의 판별을 위하여 행렬 X^tX 의 고유값이 중요한 측도라고 했다. 고유값을 상대적으로 비교하면 다중공 선성을 더 명확하게 알 수 있다. 조건지수는 가장 튼 고유값과 다른 고유값의 비율의 제곱근으로 나타내어진다.

$$\kappa_i = \sqrt{\frac{\lambda_1}{\lambda_i}}, \quad i = 2, 3, \cdots, p$$

중요한 역활을 하는 조건지수는 가장 작은 고유값에 의한 것이다.

$$\kappa_p = \kappa(\pmb{X}) = \sqrt{\frac{\lambda_1}{\lambda_p}}$$

 $\kappa(\pmb{X})$ 의 값이 크면 클수록 다중공선성의 가능성이 높다. 일반적으로 $\kappa(\pmb{X})$ 가 30 이상이면 다중공선성의 가능성이 크다고 본다.

4.3.4. 분산팽창계수 (Variance Inflation Factor; VIF)

하나의 독립변수 x_i 를 나머지 다른 p-1개의 독립변수 $x_1,\dots,x_{i-1},x_{i+1},\dots,x_p$ 를 이용하여 회귀식을 적합시킬 수 있다. 이때 다중공선성이 존재한다면 적합된 회귀식의 결정계수 R_i^2 는 1에 매우 가까울 것이다

$$x_i = \beta_0 + \beta_1 x_1 + \dots + \beta_{i-1} x_{i-1} + \beta_{i+1} x_{i+1} + \dots + \beta_p x_p + e$$

이때 독립변수 x_i 에 대한 VIF는 다음과 같이 정의되며 그 값이 5 또는 10보다 크면 다중공선성의 가능성이 크다.

$$VIF_i = \frac{1}{1 - R_i^2}$$

References

강근석, 와/과 유형조. 2016. R을 활용한 선형회귀분석. 1st ed. 교우사. https://github.com/regbook/regbook.

A. 행렬의 기초

이 장에서는 회귀분석의 이론 전개에 필요한 행렬 이론과 선형 대수의 기초에 대하여 알아볼 것이다.

A.1. 벡터와 행렬

다음 p-차원 벡터(vector) 또는 열벡터(column vector) \boldsymbol{a} 는 p개의 원소 a_1, a_2, \ldots, a_p 를 하나의 열(column)에 배치한 형태를 가진 개체이다.

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{A.1}$$

차원이 $n \times p$ 인 행렬 ${\pmb A}$ 는 다음과 같이 n개의 행과 p 개의 열에 원소 a_{ij} 를 다음과 같이 배치한 형태를 가진다.

$$\pmb{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

A.2. 두 행렬의 덧셈

두 행렬 A 와 B 를 더하는 규칙은 다음과 같다.

- 두 행렬 A 와 B 는 행과 열의 갯수가 같아야 한다.
- A + B = C 라고 하면, 덧셈의 결과로 만들어진 행렬 C는 두 행렬과 같은 수의 행과 열을 가지면 각 원소는 다음과 같다.

$$A + B = C$$
 \rightarrow $c_{ij} = a_{ij} + b_{ij}$

A.3. 스칼라곱

임의의 실수 λ (스칼라)가 주어졌을 때, λ 와 행렬 $m{A}$ 의 스칼라곱(scalar product) 는 행렬의 모든 원소에 λ 를 곱 해준 행렬로 정의된다.

예를 들어 $\lambda=2, \textbf{\textit{A}}\in\mathbb{R}^{2\times 3}$ 인 경우

$$\lambda \mathbf{A} = 2 \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ -2 & 0 & 4 \end{bmatrix}$$

A.4. 벡터와 행렬의 곱셈

 $n \times p$ 인 행렬 \boldsymbol{A} 와 p-차원 벡터(vector) \boldsymbol{b} 는 다음과 같이 두 개의 서로 다른 형태로 나타낼 수 있다.

A.4.1. 행과 열의 내적

먼저 행렬과 벡터의 곱셈은 행렬 A 의 행벡터와 벡터 b 의 내적(inner product)로 나타낼 수 있다.

$$\begin{split} \boldsymbol{A}\boldsymbol{b} &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{r}_1^t \\ \boldsymbol{r}_2^t \\ \vdots \\ \boldsymbol{r}_n^t \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} \quad \text{where } \boldsymbol{r}_i^t = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{ip} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{r}_1^t \boldsymbol{b} \\ \boldsymbol{r}_2^t \boldsymbol{b} \\ \vdots \\ \boldsymbol{r}_n^t \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^p a_{1j}b_j \\ \sum_{j=1}^p a_{2j}b_j \\ \vdots \\ \sum_{j=1}^p a_{nj}b_j \end{bmatrix} \\ &= \begin{bmatrix} < \boldsymbol{r}_1, \boldsymbol{b} > \\ < \boldsymbol{r}_2, \boldsymbol{b} > \\ \vdots \\ < \boldsymbol{r}_n, \boldsymbol{b} > \end{bmatrix} \end{aligned}$$

위에서 < a,b>는 다음과 같은 두 벡터의 내적(inner product)을 의미한다.

$$<\pmb{a},\pmb{b}>=\pmb{a}^t\pmb{b}=\sum_{i=1}^pa_ib_i$$

A.4.2. 열벡터의 선형조합

이제 행렬과 벡터의 곱셈을 행렬을 구성하는 열벡터들의 선형조합(linear combination)으로 나타낼 수 있다.

A.5. 행렬의 전치

 A^t 는 행렬의 전치(transpose)를 나타낸다. 행렬의 전치는 원소의 행과 열을 바꾸어 만든 행렬이다.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} = \{a_{ij}\}_{n \times p} \quad \rightarrow \quad \boldsymbol{A}^t = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \dots \\ a_{1p} & a_{2p} & \dots & a_{np} \end{bmatrix} = \{a_{ji}\}_{p \times n}$$

A.6. 행렬의 곱셈

먼저 두 행렬 A 와 B 의 곱셈

$$A \times B \equiv AB$$

을 정의하려면 다음과 같은 조건이 만족되어야 한다.

• 행렬 A 의 열의 갯수와 행렬 B 의 행의 갯수가 같아야 한다

따라서 두 행렬의 곱셈은 순서를 바꾸면 정의 자체가 안될 수 있다.

이제 두 행렬 $\mathbf{A} \in \mathbb{R}^{m \times n}$ 와 $\mathbf{B} \in \mathbb{R}^{n \times k}$ 의 곱셈은 다음과 같이 정의된다.

$$AB = C$$

행렬 \pmb{C} 는 m 개의 행과 k개의 열로 구성된 행렬이며($\pmb{C} \in \mathbb{R}^{m \times k}$) 각 원소 c_{ij} 는 다음과 같이 정의된다.

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lk}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, k$$

먼저 간단한 예제로 다음과 같은 두 개의 행렬의 곱을 생각해 보자.

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} (1)(0) + (2)(-1) & (1)(1) + (2)(2) \\ (3)(0) + (4)(-1) & (3)(1) + (4)(2) \end{bmatrix} = \begin{bmatrix} -2 & 5 \\ -4 & 11 \end{bmatrix}$$

곱하는 순서를 바꾸어 계산해 보자.

$$\mathbf{BA} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (0)(1) + (1)(3) & (0)(2) + (1)(4) \\ (-1)(1) + (2)(3) & (-1)(2) + (2)(4) \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$$

위 두 결과를 보면 행렬의 곱셈에서는 교환법칙이 성립하지 않음을 알 수 있다.

이제 차원이 다른 두 행렬의 곱셈을 살펴보자.

$$m{A} = egin{bmatrix} 1 & 2 & 3 \ 3 & 2 & 1 \end{bmatrix}, \quad m{B} = egin{bmatrix} 0 & 2 \ 1 & -1 \ 0 & 1 \end{bmatrix}$$

두 행렬의 곱셈은 다음과 같이 계산할 수 있다.

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix}$$

두 행렬의 곱하는 순서를 바꾸면 차원이 전혀 다른 행렬이 얻어진다.

$$\mathbf{BA} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

행렬의 곱셈은 교환법칙이 성립하지 않는다.

$$AB \neq BA$$
 (A.2)

♦ 주의

교환법칙이 성립하지 않는다는 의미는 식 A.2 이 언제나 성립한다는 의미는 아니다. 아래와 같이 특별한 경우 교환법칙이 성립하는 경우도 있다.

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$$

• 행렬의 곱셈은 결합법칙과 배분법칙은 성립한다.

$$(AB)C = A(BC)$$

$$(A+B)C = AC + BC$$

A.7. 단위벡터와 항등행렬

i번째 단위벡터 $m{e}_i$ 를 정의하자. 단위벡터 $m{e}_i$ 는 n- 차원 벡터로서 i번째 원소만 1이고 나머지는 0인 벡터이다.

$$oldsymbol{e}_i = egin{bmatrix} 0 \ 0 \ dots \ 0 \ 1 \ 0 \ dots \ 0 \end{bmatrix}$$

즉 n-차원 항등행렬 I는 n개의 단위벡터들을 모아놓은 것이다. 단위행렬은 대각원소가 1이고 나머지는 0인 정방행렬이다.

$$I = [\boldsymbol{e}_1 \ \boldsymbol{e}_2 \ \dots \ \boldsymbol{e}_n]$$

A.8. 대각합

 ${m A}=\{a_{ij}\}$ 를 $n \times n$ 정방행렬(square matrix)인 경우, 행렬의 대각 원소(diagonal element)들의 합(trace)을 $tr({m A})$ 로 표시한다.

$$tr(\pmb{A}) = \sum_{i=1}^{n} a_{ii}$$

두 행렬의 덧셈(뺄셈)에 대한 대각합에 대한 성질들은 다음과 같다.

$$tr(\boldsymbol{A} \pm \boldsymbol{B}) = tr(\boldsymbol{A}) \pm tr(\boldsymbol{B})$$

🍐 주의

행렬의 곱셈은 일반적으로 교환법칙이 성립하지 않지만 대각합의 연산은 교환법칙이 성립한다.

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

대각합은 교환법칙이 성립히기 때문에 다음과 같은 성질이 성립한다.

$$\mathrm{tr}(\pmb{AKL})=\mathrm{tr}(\pmb{KLA})$$

벡터의 연산에서도 대각합의 교환법칙이 성립되어 다음과 같은 유용한 식이 성립한다.

$$\operatorname{tr}(\boldsymbol{x}\boldsymbol{y}^t) = \operatorname{tr}(\boldsymbol{y}^t\boldsymbol{x}) = \boldsymbol{y}^t\boldsymbol{x} \in \mathbb{R}.$$

대각합의 교환법칙때문에 어떤 행렬의 앞에 특정 행렬을 곱하고, 뒤에 역행렬을 곱해도 대각합은 변하지 않는다.

$$\operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) = \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

대각합에 대한 그 밖의 성질들은 다음과 같다.

- $\operatorname{tr}(\alpha \mathbf{A}) = \alpha \operatorname{tr}(\mathbf{A}), \alpha \in \mathbb{R} \text{ for } \mathbf{A} \in \mathbb{R}^{n \times n}$
- $\operatorname{tr}(\boldsymbol{I}_n) = n$

A.9. 행렬식

 \mathbf{A} 의 행렬식(determinant)을 $det(\mathbf{A}) = |\mathbf{A}|$ 로 표기한다.

이차원 행렬 A 의 행렬식은 다음과 같이 계산한다.

$$\det(\pmb{A}) = \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| = a_{11}a_{22} - a_{12}a_{21}.$$

만약 행렬 ${m A}$ 가 대각행렬(diagonal matrix)이면 $|{m A}|$ 는 행렬의 대각원소의 곱이다 ($|{m A}|=\prod a_{ii}$). 두 행렬의 곱의 행렬식은 각 행렬의 행렬식의 곱이다.

$$|AB| = |A||B|$$

행렬식에 대한 유용한 공식들은 다음과 같다.

- $|A^t| = |A|$
- $|c\mathbf{A}| = c^n |\mathbf{A}|$

만약 행렬 A가 다음과 같은 분할행렬(partitioned matrix) 의 형태를 가지면

$$m{A} = egin{bmatrix} m{A}_{11} & m{A}_{12} \ m{0} & m{A}_{22} \end{bmatrix}$$

행렬 A의 행렬식은 다음과 같이 주어진다.

$$|A| = |A_{11}||A_{22}|$$

A.10. 직교행렬

만약 정방행렬 \boldsymbol{P} 가 다음과 같은 조건을 만족하면 직교행렬(orthogonal matrix)라고 부른다.

$$PP^t = P^tP = I$$

직교행렬의 정의에서 주의할 점은 \mathbf{SP} $\mathbf{P}^{^{*}}\mathbf{t}=\mathbf{I}$ \mathbf{S} 와 $\mathbf{P}^{t}\mathbf{P}=\mathbf{I}$ 이 모두 성립하해야 한다는 점이다. 행렬 \mathbf{P} 의 역행렬은 \mathbf{P}^{t} 이다.

$$\mathbf{P}^{-1} = \mathbf{P}^t$$

만약 P가 직교행렬이면 다음과 같은 성질을 가진다.

• $|\mathbf{P}| = \pm 1$, 왜냐하면

$$|PP^t| = |P||P^t| = |P|^2 = |I| = 1$$

• 임의의 정방행렬 A에 대하여 다음이 성립한다.

$$tr(\pmb{P}\pmb{A}\pmb{P}^t) = tr(\pmb{A}\pmb{P}^t\pmb{P}) = tr(\pmb{A})$$

B. 벡터공간

B.1. 벡터공간의 정의와 의미

먼저 지금까지 우리가 배운 벡터의 개념을 일반화하여 다루기 위해서 벡터공간의 일반적 개념을 정의하고자 한다.

벡터는 숫자를 모아놓은 형태인 식 A.1 로 주로 나타내지만 이러한 벡터를 모아놓은 집합은 실벡터 공간(real vector space)이라고 한다. 즉, 식 A.1 의 벡터는 p-차원 실벡터(real vector)라고 한다.

지금부터 논의할 추상적인 벡터 공간(abstract vector space)은 어떤 집합이든 원소에 대한 더하기와 스칼라곱이 정의되어 있는 공간을 말한다.

이제부터 $\mathbb R$ 을 실수 전체 집합이라고 하자. 또한 $\mathbb R^n$ 을 n-차원 실벡터(real vector)의 집합이라고 하자. 또한 $\mathbb R^{n\times p}$ 을 $n\times p$ -차원 행렬의 집합이라고 하자.

벡터공간(vector space) 은 어떤 집합 S 에 다음과 같은 두 개의 연산이 정의된 공간을 말한다.

1. 두 개의 원소에 대한 더하기(addition, +) 연산의 정의되어 있다.

$$+: S + S \to S$$
 (B.1)

2. 하나의 실수와 한 개의 원소에 대한 스칼라곱(scalar product, ·) 연산이 정의되어 있다.

$$\cdot : \mathbb{R} \cdot S \to S \tag{B.2}$$

위에서 더하기 연산이 정의되어 있다는 의미는 다음에 주어진 규칙이 성립한다는 의미이다.

• 집합 S 가 연산에 대하여 닫혀있다 (closure).

$$s_1 + b \in S \quad \forall s_1, b \in S$$

• 결합법칙이 성립한다 (Associativity).

$$(s_1 + s_2) + s_3 = s_1 + (s_2 + s_3) \quad \forall s_1, s_2, s_3 \in S$$

• 항등원이 존재한다 (Neutral element).

$$s+e=e+s=s \quad \exists e \ \forall s \in S$$

• 역원이 존재한다 (Inverse element).

$$s+i=i+s=0$$
 $\exists i \ \forall s \in S$

일반적으로 항등원(e) 는 0 으로 표시하며 역원(i) 는 -s 로 표시한다.

• 교환법칙이 성립한다 (Commutativity).

$$s_1+s_2=s_2+s_1 \quad \forall s_1,s_2 \in S$$

또한 위에서 스칼라곱 연산이 정의되어 있다는 의미는 다음에 주어진 규칙이 성립한다는 의미이다.

• 스칼라곱 연산의 분배법칙이 성립한다 (Distributivity).

$$r_1(s_1+s_2) = r_1s_1 + r_2s_2, \quad (r_1+r_2)s = r_1s + r_2s \quad \forall s_1, s_2 \in S, \quad \forall r_1, r_2in\mathbb{R}$$

• 스칼라곱 연산의 결합법칙이 성립한다

$$r_1(r_2s) = (r_1r_2)s \quad \forall s \in S, \ \forall r_1, r_2in\mathbb{R}$$

• 스칼라곱 연산의 항등원이 존재한다 (Neutral element).

$$1 \cdot s = s \quad \forall s \in S$$

🌢 주의

벡터 공간에서 주의할 점은 두 벡터의 곱하기 가 정의되어 있다는 것이 아니라 하나의 스칼라와 하나의 벡터에 대한 스칼라 곱하기가 정의되어 있다는 것이다.

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = ? \quad but \quad 3 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

두 벡터의 곱하기 는 내적(inner product) 란 이름으로 따로 정의한다. 또한 두 벡터의 곱셈이 유일하게 정의되지 않는다는 점에 유의하자. 예를 들어 벡터의 곱셈은 외적(cross product) 이라는 이름으로 정의된다.

이 강의에서는 스칼라로 실수만 사용하고 벡터공간은 실벡터공간(real vector space)만 고려할 것이다. 하지만 벡터공간은 실벡터가 아닌 다른 일반적인 집합에 대해서도 정의할 수 있음을 유의하자. 예를 들어 *n*-차원 다힝식들을 모두 모아 놓은 집합은 벡터공간이다. 또한 연속인 함수들을 모아 놓은 집합도 벡터공간이다.

B.2. 벡터의 선형독립

벡터공간에 속한 벡터 $m{v}_1,\ m{v}_2,\ \dots\ , m{v}_n$ 의 선형결합(또는 선형결합, linear combination)이란 각 벡터에 스칼라를 곱하여 더한 것들이다.

즉 다음과 같은 형태의 식을 벡터 $m{v}_1, \ m{v}_2, \ \dots \ , m{v}_n$ 의 선형결합(linear combination)이라고 한다:

$$r_1 \boldsymbol{v}_1 + r_2 \boldsymbol{v}_2 + \dots + r_n \boldsymbol{v}_n, \quad r_1, r_2, \dots, r_n \in \mathbb{R} \tag{B.3}$$

정의 $\mathbf{B.1}$ (벡터의 선형독립과 선형종속). 벡터공간에 속한 벡터 $\mathbf{v}_1,\ \mathbf{v}_2,\ \dots\ ,\mathbf{v}_n$ 가 있다고 하자. 만약 다음 식이 만약 모두 0인 n개의 스칼라 x_1,x_2,\dots,x_n 에 대해서만 성립하면 n개 벡터 $\mathbf{v}_1,\ \mathbf{v}_2,\ \dots\ ,\mathbf{v}_n$ 들은 선형독립 (linearly independent)라고 한다.

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \dots + x_n \mathbf{v}_n = \mathbf{0} \quad \Longleftrightarrow x_1 = x_2 = \dots = x_n = 0 \tag{B.4}$$

또한 벡터 $\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \dots\ ,\boldsymbol{v}_n$ 가 선형독립이 아니면 선형종속(linear dependent)라고 한다. 벡터 $\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \dots\ ,\boldsymbol{v}_n$ 가 선형종속이면 모두 $\boldsymbol{0}$ 이 아닌 x_1,x_2,\dots,x_n 이 존재하여 다음이 성립한다는 것이다.

$$\exists x_1, x_2, \dots, x_n \in \mathbb{R} \text{ s.t. } (x_1, x_2, \dots, x_n) \neq \mathbf{0}, \quad \mathbf{v}_1 + x_2 \mathbf{v}_2 + \dots + x_n \mathbf{v}_n = \mathbf{0}$$
 (B.5)

예를 들어 다음과 같이 주어진 3개의 3-차원 벡터들은 선형종속이다.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$$
 (B.6)

왜냐하면 다음과 같이 모두 0이 아닌 스칼라에 의해서 다음 식이 성립하기 때문이다. 즉 벡터 \pmb{v}_3 는 \pmb{v}_2 에 2를 곱하여 \pmb{v}_1 에 더한 값과 같다.

$$\mathbf{v}_3 = \mathbf{v}_1 + 2\mathbf{v}_2 \quad \Longleftrightarrow \quad \mathbf{v}_1 + 2\mathbf{v}_2 - \mathbf{v}_3 = 0$$

이제 다음과 같이 주어진 3개의 3-차원 벡터들은 선형독립이다. 즉 3개 벡터의 선형 조합이 0이 될 수 있도록 만드는 스칼라는 모두 0인 경우 밖에 없다.

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$$
 (B.7)

이제 다음과 같이 주어진 4개의 3-차원 벡터들은 선형종속이다.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$
 (B.8)

 $oldsymbol{v}_3$ 가 다음과 같이 다른 벡터의 선형결합으로 나타난는 것을 보여준다.

$$\mathbf{\textit{v}}_{3} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = (1)\mathbf{\textit{v}}_{1} + (2)\mathbf{\textit{v}}_{2} + (-1)\mathbf{\textit{v}}_{4} = (1)\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + (2)\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

식 B.8 와 같이 3차원 벡터가 4개인 경우 벡터의 값에 관계없이 선형종속으로 나타난다. 이러한 사실은 \mathbb{R}^n 의 n+1개의 벡터는 항상 선형종속이라는 정리의 결과이다.

즉, \mathbb{R}^n 에서 n 개보다 더 많은 벡터들은 항상 선형종속이다.

B.3. 역행렬

정방행렬 A 의 역행렬(inverse matrix) A^{-1} 는 다음과 같은 성질을 만족하는 행렬이다.

$$AA^{-1} = A^{-1}A = I$$

역행렬은 언제나 존재하는 것은 아니다. 만약 행렬 \pmb{A} 가 역행렬을 가지면 이를 \pmb{A}^{-1} 로 표시한다. 또한 역행렬이 존재하면 정칙행렬(non-singular matrix)이라고 한다.

역행렬은 존재하는 조건은 행렬식(determinant)이 0이 아니어야 한다.

B.4. 행렬의 계수

행렬의 계수(rank)란 일차 독립인 열들의 최대 수 또는 일차 독립인 행들의 최대 수로 정의된다

$$rank(\mathbf{A}) = rk(\mathbf{A}) = dim(Col(\mathbf{A})) = dim(Row(\mathbf{A}))$$

꼭 기억해야 할 것은 행렬의 계수는 열들을 이용하여 구한 계수와 행들을 이용하여 구한 계수가 같다는 것이다. 즉, 행렬의 계수는 열의 계수와 행의 계수 중 하나만 구해도 된다는 것이다.

예를 들어 식 B.6 에 주어진 3 개의 벡터를 열로 하는 행렬의 계수는 2이다. 왜냐하면 선형종속인 벡터가 하나 있기 때문이다.

$$m{A} = egin{bmatrix} 1 & 1 & 3 \\ 2 & 0 & 2 \\ 3 & 1 & 5 \end{bmatrix} \quad o \quad rank(m{A}) = 2$$

위에 주어진 행렬 A의 행들을 고려하면 첫 번째 행과 두 번째 행의 합이 세 번째 행으로 나타난다. 즉, 서로 독립인 행의 최대 개수는 2 이며 이는 서로 독립인 열의 최대 개수와 같다. 따라서 행렬 A의 계수는 2이다.

다음으로 식 B.7 에 주어진 3 개의 벡터를 열로 하는 행렬의 계수는 3이다. 3개의 열벡터와 3개의 행벡터들은 모두 선형독립이다.

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 0 & 2 \\ 3 & 1 & 4 \end{bmatrix} \rightarrow rank(A) = 3$$

이제 식 B.8 주어진 4개의 벡터로 이루어진 행렬의 계수는 3이다. 왜냐하면 4개의 열벡터 중 3개의 열벡터는 선형독립이지만 4번째 열벡터는 선형종속이기 때문이다.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 3 & 0 \\ 2 & 0 & 2 & 0 \\ 3 & 1 & 4 & 1 \end{bmatrix} \rightarrow rank(\mathbf{A}) = 3$$

행렬의 열과 행의 개수가 다를 때 행렬의 계수는 열의 개수와 행의 개수 중 작은 값보다 같거나 작다 예를 들어 $n \times p$ 행렬의 계수는 min(n,p) 와 같거나 작다.

$$A \in \mathbb{R}^{n \times p} \quad \rightarrow \quad rank(\mathbf{A}) \leq min(n, p)$$

다음은 행렬의 계수에 관련된 주요 공식이다.

- $rank(\mathbf{A}) = rank(\mathbf{A}^t)$
- 행렬 \boldsymbol{A} 가 정방행렬이고 계수가 n 이면 역행렬이 존재한다(정칙행렬).
- 또한 더 나아가 A 가 정칙행렬이라는 사실은 아래 나열된 조건들과 동치(equivalance)이다.
 - $\Leftrightarrow A$ 의 열들이 일차독립이다.
 - $\Leftrightarrow A$ 의 행들이 일차독립이다.
 - $\Leftrightarrow \mathbf{A}$ 의 계수가 n 이다.
 - $\Leftrightarrow A$ 의 행렬식이 0이 아니다.

B.5. 생성집합과 기저

벡터공간 V 의 벡터 $oldsymbol{v}_1,oldsymbol{v}_n,\dots,oldsymbol{v}_m$ 의 선형결합을 모두 모은 집합

$$W = span\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\} = \{r_1\boldsymbol{v}_1 + r_2\boldsymbol{v}_2 + \dots + r_m\boldsymbol{v}_m : r_1, r_2, \dots, r_m \in \mathbb{R}\}$$

을 벡터 $\pmb{v}_1, \pmb{v}_n, \dots, \pmb{v}_m$ 의 생성(span)이라고 하며 W 의 생성집합(generating set, spanning set) 이라고 한다.

또한 어떤 벡터공간의 생성집합에 속한 벡터들이 선형독립일 때 이 생성집합을 기저 (basis)라고 한다.

만약 주어진 벡터 공간의 부분집합이 다시 벡터공간의 정의를 만족한다면 이를 부분공간(subspace)이라고 한다. 위에서 정의한 생성집합 W는 벡터공간 V의 부분공간이다.

B.6. 벡터공간의 차원

- \mathbb{R}^n 의 모든 기저는 n개의 원소를 갖는다.
- 임의의 벡터공간 V에 대해서 V의 부분집합 $B=\{m{b}_1,\dots,m{b}_n\}$ 가 V의 기저라고 하면 다음을 보일 수 있다.
 - -V 의 모든 벡터들은 $\boldsymbol{b}_1,\dots,\boldsymbol{b}_n$ 의 선형결합으로 나타낼 수 있으며 유일하다.
 - -V 의 부분집합이 n 개보다 많은 벡터를 포함하면 이 부분집합의 벡터들은 선형종속이다.
 - V 의 또 다른 기저 $C = \{c_1, ..., c_m\}$ 가있다면 m = n 이다.
- 벡터공간 V의 차원(dimension) 은 기저의 개수로 정의되며 dim(V)로 표시한다.

B.7. 행렬의 열공간과 행공간

 $n \times p$ 행렬 \pmb{A} 에 의하여 생성되는 열공간(column space) $C(\pmb{A})$ 는 행렬 \pmb{A} 를 구성하는 열벡터의 선형조합으로 나타낼 수 있는 모든 벡터들의 집합을 말한다.

$$C(\mathbf{A}) = \{ \mathbf{y} | \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^p \} \subset \mathbb{R}^n$$

 $n \times p$ 행렬 \boldsymbol{A} 에 의하여 생성되는 영공간(null space) $N(\boldsymbol{A})$ 는 다음과 같이 정의되는 벡터들의 집합을 말한다.

$$N(\mathbf{A}) = \{ \mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{0} \text{ for } \mathbf{x} \in \mathbb{R}^p \} \subset \mathbb{R}^n$$

벡터공간과 영공간은 다음과 같은 성질을 가진다.

- $rank(\mathbf{A}) = dimension of C(\mathbf{A}) = dim[C(\mathbf{A})]$
- $dim[C(\mathbf{A})] + dim[N(\mathbf{A})] = n$

B.8. 두 벡터의 사영

선형독립인 두 벡터 \pmb{a}_1 과 \pmb{a}_2 가 있다고 하자. 벡터 \pmb{a}_1 과 같은 방향을 가지면서 벡터 \pmb{a}_2 에 가장 가까운 벡터를 $proj_{\pmb{a}_1}(\pmb{a}_2)$ 라고 정의하고 이를 벡터 \pmb{a}_1 방향으로 벡터 \pmb{a}_2 의 사영(projection)이라고 부른다.

그러면 이러한 사영은 어떻게 구할 수 있나? 벡터 ${m a}_2$ 의 사영은 벡터 ${m a}_1$ 방향에 있으므로 어떤 실수 c 가 있어서 다음과 같이 표시할 수 있다.

$$proj_{\boldsymbol{a}_1}(\boldsymbol{a}_2) = c\boldsymbol{a}_1$$

이제 사영 $c \boldsymbol{a}_1$ 과 벡터 \boldsymbol{a}_2 의 거리 d(c) 를 생각하면 다음과 같다.

$$\begin{split} d^2(c) &= \left\| \mathbf{a}_2 - c \mathbf{a}_1 \right\|^2 \\ &= (\mathbf{a}_2 - c \mathbf{a}_1)^t (\mathbf{a}_2 - c \mathbf{a}_1) \\ &= \mathbf{a}_2^t \mathbf{a}_2 - 2c \mathbf{a}_2^t \mathbf{a}_1 + c^2 \mathbf{a}_1^t \mathbf{a}_1 \end{split}$$

위에서 $\|a\|$ 는 벡터 a의 길이를 나타낸다.

$$d(\boldsymbol{a}) = \|\boldsymbol{a}\| = \sqrt{\boldsymbol{a}^t \boldsymbol{a}}$$

상수 c 는 거리 d(c)를 최소로 만드는 수이다. $d^2(c)$ 은 c 에 대하여 미분 가능한 2차 함수이며 아래로 볼록한 함수 이므로 이를 미분하여 c 를 구할 수 있다.

$$\frac{\partial d^2(c)}{\partial x}c = -2\boldsymbol{a}_2^t\boldsymbol{a}_1 + 2c\boldsymbol{a}_1^t\boldsymbol{a}_1 = 0$$

위의 방적식으로 부터 c를 얻고

$$c = \frac{\boldsymbol{a}_2^t \boldsymbol{a}_1}{\boldsymbol{a}_1^t \boldsymbol{a}_1}$$

다음과 같이 벡터 \boldsymbol{a}_1 방향으로 벡터 \boldsymbol{a}_2 의 사영을 나타낼 수 있다.

$$proj_{\boldsymbol{a}_1}(\boldsymbol{a}_2) = \frac{\boldsymbol{a}_2^t \boldsymbol{a}_1}{\boldsymbol{a}_1^t \boldsymbol{a}_1} \boldsymbol{a}_1 \tag{B.9}$$

이제 위의 두 벡터의 사영을 이용하면 벡터 $m{a}_1$ 과 직교하는 벡터 $m{ ilde{q}}_2$ 를 다음과 같이 찾을 수 있다.

$$\tilde{\boldsymbol{q}}_2 = \boldsymbol{a}_2 - proj_{\boldsymbol{a}_1}(\boldsymbol{a}_2) = \boldsymbol{a}_2 - \frac{\boldsymbol{a}_2^t \boldsymbol{a}_1}{\boldsymbol{a}_1^t \boldsymbol{a}_1} \boldsymbol{a}_1$$

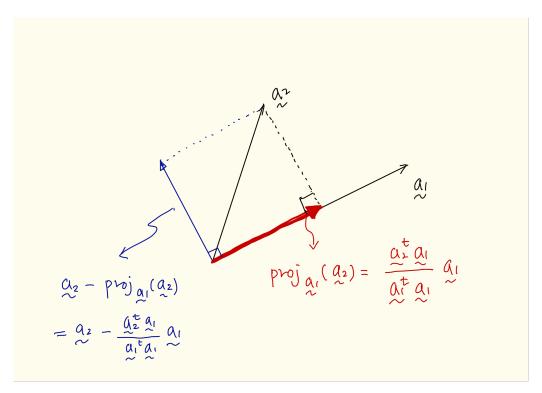


그림 B.1.: 벡터의 사영

두 벡터 \pmb{a}_1 와 $\tilde{\pmb{q}}_2$ 의 직교성은 다음과 같이 보일 수 있다.

$$egin{aligned} m{a}_1^t & m{ ilde{q}}_2 = m{a}_1^t \left(m{a}_2 - rac{m{a}_2^t m{a}_1}{m{a}_1^t m{a}_1} m{a}_1
ight) \ & = m{a}_1^t m{a}_2 - rac{m{a}_2^t m{a}_1}{m{a}_1^t m{a}_1} m{a}_1^t m{a}_1 \ & = m{a}_1^t m{a}_2 - m{a}_2^t m{a}_1 \ & = 0 \end{aligned}$$

이제 두 벡터 ${m q}_1$ 과 ${m q}_2$ 를 다음과 같이 정규직교벡터로 만들 수 있다.

$$\begin{split} & \boldsymbol{q}_1 = \boldsymbol{a}_1 / \left\| \boldsymbol{a}_1 \right\| \\ & \boldsymbol{q}_2 = \tilde{\boldsymbol{q}}_2 / \left\| \tilde{\boldsymbol{q}}_2 \right\| \end{split}$$

B.9. 최소제곱법과 사영

회귀계수벡터의 값을 구하는 최소제곱법의 기준을 다시 살펴보자.

$$\min_{\pmb{\beta}}(\pmb{y}-\pmb{X}\pmb{\beta})^t(\pmb{y}-\pmb{X}\pmb{\beta})$$

위에서 $oldsymbol{X}oldsymbol{eta}$ 는 행렬 $oldsymbol{X}$ 의 열벡터 $oldsymbol{x}_1,oldsymbol{x}_2,\dots,oldsymbol{x}_p$ 로 이루어진 선형조합이다.

$$\boldsymbol{X}\boldsymbol{\beta} = [\boldsymbol{x}_1 \ \cdots \boldsymbol{x}_p]\boldsymbol{\beta} = \beta_1 \boldsymbol{x}_1 + \cdots + \beta_p \boldsymbol{x}_p$$

행렬 \boldsymbol{X} 의 열벡터로 생성된 공간을 $C(\boldsymbol{X})$ 라고 하자

$$C(\pmb{X}) = span\{\pmb{x}_1, \dots, \pmb{x}_p\} = \{\pmb{X}\pmb{\beta} \mid \pmb{\beta} \in \mathbb{R}^p\}$$

따라서 최소제곱법으로 구한 회귀계수 벡터 $\hat{oldsymbol{eta}}$ 는 반응값 벡터 $oldsymbol{y}$ 와 $oldsymbol{X}oldsymbol{eta}$ 의 거리가 최소가 되도록 만들어 준다.

$$\begin{split} \min_{\pmb{\beta}} (\pmb{y} - \pmb{X} \pmb{\beta})^t (\pmb{y} - \pmb{X} \pmb{\beta}) &= (\pmb{y} - \pmb{X} \hat{\pmb{\beta}})^t (\pmb{y} - \pmb{X} \hat{\pmb{\beta}}) \\ \hat{\pmb{\beta}} &= (\pmb{X}^t \pmb{X})^{-1} \pmb{X}^t \pmb{y} \end{split}$$

따라서 예측값 벡터 $\hat{\boldsymbol{y}}$ 는 행렬 \boldsymbol{X} 의 열벡터로 생성한 열공간 $C(\boldsymbol{X})$ 방향으로 반응값 벡터 \boldsymbol{y} 를 사영한 벡터이다.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

위에서 행렬 $X(X^tX)^{-1}X^t$ 를 열공간 C(X)의 사영행렬(projection matrix)라고 부른다. 사영행렬의 정의는 부록 F 에서 공부한다.

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t \tag{B.10}$$

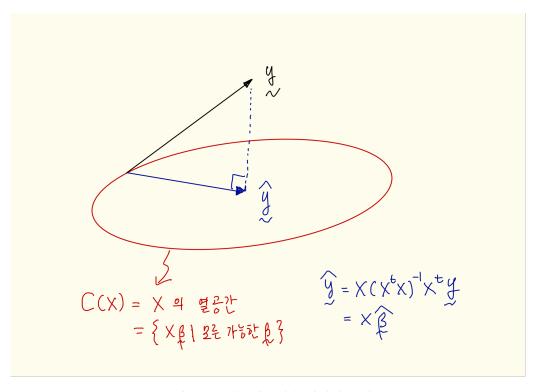


그림 B.2.: 최소제곱법을 설명한 그림

C. 고유값과 고유벡터

💧 주의

고유값과 고유벡터는 정방행렬(square matrix)에 대해서만 정의된다.

C.1. 특성다항식

특성다항식(Characteristic polynomial)은 다음과 같이 정의된다

실수 $\lambda \in \mathbb{R}$ 와 정방행렬(square matrix) $A \in \mathbb{R}^{n \times n}$ 에 대하여

$$\begin{split} p_A(\lambda) &:= \det(A - \lambda I) \\ &= c_0 + c_1 \lambda + c_2 \lambda^2 + \dots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n, \end{split} \tag{C.1}$$

C.2. 고유값과 고유벡터

C.2.1. 정의

n-차원 정방행렬 \boldsymbol{A} 이 있을 때, 다음 식을 만족하는 λ 와 벡터 \boldsymbol{x} 가 존재하면 λ 를 행렬 \boldsymbol{A} 의 고유값(eigenvalue), \boldsymbol{x} 를 행렬 \boldsymbol{A} 의 고유벡터(eigenvector)라고 한다 (부교재 definition 4.6)

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

• 고유벡터는 유일하지 않다. 즉, 벡터 x 가 고유벡터이면 cx 도 고유벡터이다.

$$A(c\mathbf{x}) = cA\mathbf{x} = c\lambda\mathbf{x} = \lambda(c\mathbf{x})$$

C.2.2. 계산

다음 3개의 문장은 동치이다

- λ 는 행렬 \boldsymbol{A} 의 고유값이다.
- 방정식 $(\pmb{A} \lambda \pmb{I}) \pmb{x} = \pmb{0}$ 은 영벡터이외의 해를 가진다(nontrivial solution)
- λ 는 행렬 $A \lambda I$ 의 행렬식이 0이다.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \tag{C.2}$$

- $\lambda \vdash \text{will} \mathbf{A} \lambda \mathbf{I}$ 의 rank가 n 보다 작다.
- Theorem 4.8 에 의하면 위에서 행렬식이 0 인 방정식 식 C.2 을 푸는 것은 식 C.1 의 이 0 을 푸는 것과 동 일하다는 것이다.

C.2.3. 중복도와 고유공간

- 대수적 중복도(algebraic multiplicity) 는 특성다항식 4 C.1 이 0인 방정식을 푸는 경우 다항식에서 고유 값이 중근(multiple root)의 해로 나타나는 차수를 의미한다.
- 기하적 중복도(geometric multiplicity) 는 고유값에 대응하는 고유벡터들 중 선형독립인 고유벡터들의 최 대 개수를 의미한다.
- 고유 공간(eigenspace)은 고유값에 대응하는 고유벡터들이 생성하는 벡터공간을 의미한다.

예제 C.1. 3차원 행렬 A 가 다음과 같을 때

$$\mathbf{A} = \left[\begin{array}{rrr} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{array} \right]$$

행렬 A의 특성다항식은 다음과 같다.

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2$$

참고로 특성방정식을 푸는 경우, 방정식 $\det(\pmb{A}-\lambda \pmb{I})=0$ 이나 $\det(\lambda \pmb{I}-\pmb{A})=0$ 중 어느 것을 사용해도 상관없다. 첫번째 고유값은 $\lambda_1=1$ 이다. 고유벡터를 구하기 위해서는 다음과 같은 방정식을 풀면 된다.

$$(\lambda_1 \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$

위의 방정식을 풀면

$$(\lambda_1 \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

아래와 같이 간단히 할 수 있으며

$$x_1 = -2x_3, \quad x_2 = x_3$$

다음과 같은 고유값과 고유벡터를 얻을 수 있다.

$$\lambda_1 = 1 \quad o \quad \boldsymbol{x}_1 = egin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

첫번째 고유값은 $\lambda_1=1$ 이며 대수적 중복도는 1이고 기하적 중복도도 1이다. 이 경우 고유공간 E_1 은 한 개의 고유벡터 $\textbf{\textit{x}}_1$ 이 생성하는 부분공간을 의미한다.

$$E_1 = \operatorname{span} \left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\}$$

다음으로 두번째 고유값에 대한 방정식 $(\lambda_2 I - A)x = 0$ 을 풀면 다음과 같다.

$$(\lambda_2 \pmb{I} - \pmb{A}) \pmb{x} = (2\pmb{I} - \pmb{A}) \pmb{x} = \begin{bmatrix} 2 & 0 & 2 \\ -1 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

이 방정식은 아래와 같이 간단히 할 수 있으며

$$x_1 = -x_3$$

다음과 같은 두 개의 고유벡터를 얻을 수 있다.

$$egin{aligned} egin{aligned} oldsymbol{\lambda}_2 = 2 &
ightarrow & oldsymbol{x}_2 = egin{bmatrix} -1 \ 0 \ 1 \end{bmatrix} & oldsymbol{x}_3 = egin{bmatrix} 0 \ 1 \ 0 \end{bmatrix} \end{aligned}$$

위에서 두번째 고유값은 $\lambda_2=2$ 이며 대수적 중복도는 ${f 2}$ 이다. 또한 선형독립인 ${f 2}$ 개의 고유벡터를 구할 수 있으므로 기하적 중복도는 ${f 2}$ 이다.

이 경우 E_2 는 두 개의 고유벡터 \pmb{x}_2, \pmb{x}_3 가 생성하는 부분공간을 의미한다.

$$E_2 = \operatorname{span} \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

이제 대수적 중복도와 기하적 중복도가 다른 경우에 대한 예제를 들어보자.

예제 C.2. 3차원 행렬 A 가 다음과 같을 때

$$\mathbf{A} = \left[\begin{array}{rrr} 1 & 0 & 2 \\ -1 & 1 & 3 \\ 0 & 0 & 2 \end{array} \right]$$

행렬 A의 특성다항식은 다음과 같다.

$$\det(\lambda \pmb{I} - \pmb{A}) = \left| \begin{array}{ccc} \lambda - 1 & 0 & -2 \\ 1 & \lambda - 1 & -3 \\ 0 & 0 & \lambda - 2 \end{array} \right| = (\lambda - 1)^2 (\lambda - 2)$$

첫번째 고유값은 $\lambda_1=1$ 이다. 고유벡터를 구하기 위해서는 다음과 같은 방정식을 풀면 된다.

$$(\lambda_1 \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$

위의 방정식을 풀면

$$(\lambda_1 \mathbf{I} - \mathbf{A})\mathbf{x} = (\mathbf{I} - \mathbf{A})\mathbf{x} = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 0 & -3 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

아래와 같이 간단히 할 수 있으며

$$x_1 = x_3 = 0$$

다음과 같은 하나의 고유벡터를 얻을 수 있다.

C. 고유값과 고유벡터

$$\lambda_1 = 1 \quad \rightarrow \quad x_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

첫번째 고유값은 $\lambda_1=1$ 이며 대수적 중복도는 2이지만 기하적 중복도는 1이다. 이 경우 고유공간 E_1 은 한 개의 고유벡터 \pmb{x}_1 이 생성하는 부분공간을 의미한다.

$$E_1 = \operatorname{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

다음으로 두번째 고유값에 대한 방정식 $(\lambda_2 I - A)x = 0$ 을 풀면 다음과 같다.

$$(\lambda_2 \boldsymbol{I} - \boldsymbol{A}) \boldsymbol{x} = (2\boldsymbol{I} - \boldsymbol{A}) \boldsymbol{x} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

이 방정식은 아래와 같이 간단히 할 수 있으며

$$x_1 = -2x_3, \quad x_2 = 5x_3$$

다음과 같은 한 개의 고유벡터를 얻을 수 있다.

$$\lambda_2 = 2 \quad \rightarrow \quad x_2 = \begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix}$$

위에서 두번째 고유값은 $\lambda_2=2$ 이며 대수적 중복도는 1이다. 또한 선형독립인 1개의 고유벡터를 구할 수 있으므로 기하적 중복도는 1이다.

이 경우 E_2 는 한 개의 고유벡터 \pmb{x}_2 가 생성하는 부분공간을 의미한다.

$$E_2 = \operatorname{span} \left\{ \begin{bmatrix} -2\\5\\1 \end{bmatrix} \right\}$$

C.3. 대칭행렬의 대각화

n차원 대칭행렬 \boldsymbol{A} 에 대하여 직교행렬 \boldsymbol{P} 가 존재하여 다음과 같은 분해가 가능하다.

$$\mathbf{P}^{t}\mathbf{A}\mathbf{P} = \mathbf{\Lambda} = diag(\lambda_{1}, \lambda_{2}, \dots, \lambda_{n})$$
 (C.3)

식 $\mathbf{F}.2$ 의 분해에서 λ_i 는 행렬 $m{A}$ 의 고유치이며 행렬 $m{P}$ 의 i 번째 열은 대응하는 고유벡터 $m{p}_i$ 로 구성되어 있다.

$$\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_n]$$

이제 위의 분해를 증명해 보자. 고유치 λ_i 와 대응하는 고유벡터 ${\pmb p}_i$ 의 정의에 따라서 다음과 같은 n개의 식을 얻을 수 있고

$$Ap_i = \lambda_i p_i, \quad i = 1, 2, 3 \dots, n$$

위의 식을을 합쳐서 표기하면 다음과 같은 식을 얻으며 이는 식 F.2 를 의미한다.

$$AP = P\Lambda$$

식 F.2 를 다시 쓰면 다음과 같은 스펙트럴 분해(spectral decomposition)를 얻는다.

$$\boldsymbol{A} = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^t = \sum_{i=1}^n \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i^t$$
 (C.4)

참고로 다음의 유용한 두 식을 기억하자.

$$tr(\pmb{A}) = \sum_i \lambda_i, \quad |\pmb{A}| = \prod_i \lambda_i$$

대칭행렬의 분해 (ref?)(eq:symmdecomp1)를 이용하면 다음과 같은 이차형식의 분해를 얻을 수 있다.

$$Q(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^t \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^t \boldsymbol{x} = \boldsymbol{y}^t \boldsymbol{\Lambda} \boldsymbol{y} = \sum_{i=1}^n \lambda_i y_i^2 (\#eq : quaddecomp)$$
 (C.5)

이차형식의 분해식 @ref(eq:quaddecomp) 를 보면 행렬 \boldsymbol{A} 의 모든 고유치가 0보다 크면 양정치 임을 알 수 있다. 또한 모든 고유치가 0보다 크거나 같으면 양반정치 임을 알 수 있다.

또한 $rank(\mathbf{A}) = rank(\mathbf{\Lambda})$ 이며 이는 0이 아닌 고유치의 개수가 행렬 \mathbf{A} 의 계수(rank)임을 알 수 있다.

D. 벡터 미분

D.1. 용어

• vector differential: 벡터 미분

• partial derivative: 편미분

• gradient: 그레디언트

• Jacobian: 야코비안, 자코비안

D.2. 벡터 미분의 표기법

이제 다변량함수(multivariate function), $f: \mathbb{R}^n \to \mathbb{R}^m$ 에 대한 미분을 생각해보자.

먼저 간단한 예제를 고려해 보자. 두 열벡터

$$m{x} = egin{bmatrix} x_1 \ x_2 \end{bmatrix} \in \mathbb{R}_2, \quad m{y} = egin{bmatrix} y_1 \ y_2 \ y_3 \end{bmatrix} \in \mathbb{R}^3$$

를 고려하고 다음과 같은 함수로 두 벡터의 관계가 정의된다고 하자.

$$y_1 = x_1^2 + x_2, \quad y_2 = \exp(x_1) + 3x_2, \quad y_3 = \sin(x_1) + x_2^3 \tag{D.1} \label{eq:D.1}$$

위의 관계를 함수 관계 $\mathbf{f}: \mathbb{R}^2 \to \mathbb{R}^3$ 로 나타내보면

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ f_3(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2 \\ \exp(x_1) + 3x_2 \\ \sin(x_1) + x_2^3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \boldsymbol{y}$$

이러한 경우 다변량 함수 f를 벡터 x로 미분하려면, 즉 미분 표기법을 이용하려면 편미분을 한 결과를 행렬의 형태를 정해야한다.

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = (n \times m) - \text{matrix} \quad \text{or} \quad (m \times n) - \text{matrix}$$
?

일단 각각의 편미분 $\frac{\partial f_i}{\partial x_i}$ 를 구해야 하며 이는 scalar 미분으로 쉽게 구해진다.

$$\begin{split} \frac{\partial f_1}{\partial x_1} &= 2x_1, & \frac{\partial f_2}{\partial x_1} &= \exp(x_1), & \frac{\partial f_3}{\partial x_1} &= \cos(x_1) \\ \frac{\partial f_1}{\partial x_2} &= 1, & \frac{\partial f_2}{\partial x_1} &= 3, & \frac{\partial f_3}{\partial x_1} &= 3x_2^2 \end{split} \tag{D.2}$$

이제 이제 편미분값들을 행렬의 형태로 정리해보자. 편미분을 행렬에 배치할 떄 다음과 같은 규칙을 사용할 것 이다.

- 행렬의 행은 x의 차원 n 과 같다.
- 행렬의 열은 \boldsymbol{f} 의 차원 m 과 같다.

위와 같이 편미분을 배치하는 벡타 미분 표기법을 분모 표기법 (denominator layout)이라고 한다.

i 분모 표기법

$$\boldsymbol{J} = \nabla_x \boldsymbol{x} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \equiv \frac{\partial \boldsymbol{f}^t}{\partial \boldsymbol{x}} \stackrel{\equiv}{\equiv} \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \frac{\partial f_3}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_3}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & \exp(x_1) & \cos(x_1) \\ 1 & 3 & 3x_2^2 \end{bmatrix}$$

 $m{J}$ 는 야코비안 행렬(Jacobian matrix)이라고 부른다.

이제 이러한 분자표기법의 특별한 결과를 알아보자

• $f: \mathbb{R}^n \to \mathbb{R}^1$ 인 경우

 $f: \mathbb{R}^n \to \mathbb{R}^1$ 인 경우 벡터미분 결과를 그레디언트(gradient)라고 부르며 다음과 같이 표기된다.

$$\nabla_x f = \frac{\partial f}{\partial \pmb{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

• $f: \mathbb{R}^1 \to \mathbb{R}^m$ 인 경우

$$\frac{\partial \boldsymbol{f}}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \vdots \\ \frac{\partial f_m}{\partial x} \end{bmatrix}$$

참고로 식 D.1 에서 정의한 함수 관계를 두 벡터 x 와 y 의 사상관계로 보면

$$oldsymbol{f}:oldsymbol{x}\mapstooldsymbol{y}$$

다음과 같이 그레디언트 벡터를 표기할 수 있다.

D. 벡터 미분

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \frac{\partial y_3}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_3}{\partial x_2} \end{bmatrix}$$

D.3. 함성함수의 미분법

이제 합성함수의 미분법(chain rule)에 대하여 알아보자.

두 개의 함수

$$\boldsymbol{q}: \mathbb{R}^n \mapsto \mathbb{R}^m, \quad \boldsymbol{f}: \mathbb{R}^m \mapsto \mathbb{R}^p$$

가 있을 때, f와 g의 합성함수 h는 다음과 같이 정의된다.

$$h(x) = f(g(x)) = f \circ g$$

즉,

$$\pmb{h}: \mathbb{R}^n \mapsto \mathbb{R}^m \mapsto \mathbb{R}^p$$

이러한 합성함수의 미분은 다음과 같이 계산된다.

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \frac{\partial \mathbf{f} \circ \mathbf{g}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \frac{\partial \mathbf{f}}{\partial \mathbf{g}}$$
(D.3)

식 D.3 에서 $\frac{\partial \mathbf{f}}{\partial \mathbf{g}}$ 는 $m \times p$ Jacovian 벡터이고

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{g}} \begin{bmatrix} \frac{\partial f_1}{\partial g_1} & \frac{\partial f_2}{\partial g_1} & \cdots & \frac{\partial f_p}{\partial g_1} \\ \frac{\partial f_1}{\partial g_2} & \frac{\partial f_2}{\partial g_2} & \cdots & \frac{\partial f_p}{\partial g_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial g_m} & \frac{\partial f_2}{\partial g_m} & \cdots & \frac{\partial f_p}{\partial g_m} \end{bmatrix} = (m \times p)$$

 $\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}$ 는 $n \times m$ Jacovian 벡터이다

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \frac{\partial g_2}{\partial x_n} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix} = (n \times m)$$

함성함수의 미분 공식을 차원으로 나타내면 다음과 같다.

D. 벡터 미분

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{g}}$$

$$\underset{n \times m}{n \times m} \underset{m \times p}{\boldsymbol{m} \times p}$$

D.4. 두 벡터 내적의 미분

D.4.1. 상수벡터와 변수벡터의 내적

먼저 상수 벡터 a와 변수 벡터 x의 내적의 미분을 생각해 보자.

참고로 다음과 같이 두 벡터의 내적은 스칼라이다.

$$\pmb{a}^t\pmb{x} = \pmb{x}^t\pmb{a} = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

따라서 그레이디언트를 구하는 방법과 같이 결과는 열벡터로 표기된다.

$$rac{\partial oldsymbol{a}^t oldsymbol{x}}{\partial oldsymbol{x}} = rac{\partial oldsymbol{x}^t oldsymbol{a}}{\partial oldsymbol{x}} = oldsymbol{a} = egin{bmatrix} a_1 \ a_2 \ dots \ a_n \end{bmatrix}$$

위의 식에서 상수벡터 a는 가 전치로 앞에 나타나는 표현 x^ta 를 사용하면 결과 벡터 a가 열벡터로 그대로 나타나 지므로 내적의 미분 표기로 사용할 것이다.

$$\frac{\partial \mathbf{x}^t \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^t}{\partial \mathbf{x}} \mathbf{a} = I\mathbf{a} = \mathbf{a}$$
 (D.4)

D.4.2. 상수벡터와 함수벡터의 내적

더 나아가서 상수 벡터 a와 함수 벡터 f의 내적의 미분도 식 D.4을 표시하는 동일한 논리로 다음과 같이 표기할 수 있다.

$$\frac{\partial \mathbf{f}^t \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}^t}{\partial \mathbf{x}} \mathbf{a} \tag{D.5}$$

참고로 식 $\mathrm{D.5}$ 에서 $\frac{\partial f}{\partial x}$ 는 행벡터가 아닌 행렬로 나타날 수 있다.

D.4.3. 함수벡터와 함수벡터의 내적

이제 다음과 같이 같은 공간으로 사상되는 두 함수 f 와 g 의 내적을 생각해 보자.

$$f: \mathbb{R}^n \mapsto \mathbb{R}^m, \quad g: \mathbb{R}^n \mapsto \mathbb{R}^m$$

두 함수의 내적을 미분하는 경우 곱셉 법칙을 적용하여야 하는데 행렬의 곱셉에서는 교환법칙이 성립되지 않으므로 순서에 주의해야 한다.

내적 $m{f}^tm{g}$ 를 각각 따로 미분해야 하는데 각 벡터에 대해 따로 미분을 실행해 보자

• f 를 미분하는 경우 q 는 상수 벡터 a 로 취급한다. 그리고 식 D.5 를 적용한다.

$$\frac{\partial \mathbf{f}^t \mathbf{g}}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}^t \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}^t}{\partial \mathbf{x}} \mathbf{a} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{g}$$
(D.6)

• g 를 미분하는 경우 f 는 상수 벡터 a 로 취급한다. 그리고 식 $\mathrm{D.5}$ 를 적용한다.

$$\frac{\partial \mathbf{f}^t \mathbf{g}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^t \mathbf{g}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^t \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^t}{\partial \mathbf{x}} \mathbf{a} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{f}$$
 (D.7)

이제 위의 두 결과 식 D.6 과 식 D.7 를 합치면 다음과 같은 최종적인 결과를 얻을 수 있다.

$$\frac{\partial \mathbf{f}^t \mathbf{g}}{\partial \mathbf{r}} = \frac{\partial \mathbf{f}}{\partial \mathbf{r}} \mathbf{g} + \frac{\partial \mathbf{g}}{\partial \mathbf{r}} \mathbf{f}$$
 (D.8)

D.5. 벡터 미분의 응용

D.5.1. 선형사상의 미분상

이제 앞에서 배운 벡터의 미분을 이용하여 유용한 응용 공식을 유도해보자.

먼저 선형변환 $m{y} = m{A}m{x}$ 를 생각해 보자. 이때 $(M \times N) - m{A}$ 는 상수 행렬이다. 이때 $m{y}$ 를 $m{x}$ 로 미분하면 다음과 같다. 먼저 행렬 $m{A}$ 의 i 번째 행을 $m{a}_i^t$ 라고 하면

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & A_{22} & \dots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \dots & A_{MN} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^t \\ \boldsymbol{a}_2^t \\ \vdots \\ \boldsymbol{a}_M^t \end{bmatrix}$$

선형변환 f(x) = Ax 로 정의하면 다음과 같이 나타낼 수 있다.

D. 벡터 미분

$$egin{aligned} m{A}m{x} = egin{bmatrix} m{a}_1^tm{x} \ m{a}_2^tm{x} \ dots \ m{a}_M^tm{x} \end{bmatrix} = egin{bmatrix} f_1(m{x}) \ f_2(m{x}) \ dots \ f_M(m{x}) \end{bmatrix} = m{f}(m{x}) \end{aligned}$$

따라서

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_M}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_N} & \frac{\partial f_2}{\partial x_N} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{M1} \\ A_{12} & A_{22} & \cdots & A_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1N} & A_{2N} & \cdots & A_{MN} \end{bmatrix} = \boldsymbol{A}^t$$

따라서 선형사상의 미분은 선형변환 행렬의 전치이다.

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^t \tag{D.9}$$

E. 다변량 확률변수의 성질

E.1. 일변량분포

일변량 확률변수 X가 확률밀도함수 f(x)를 가지는 분포를 따를때 기대값과 분산은 다음과 같이 정의된다.

$$E(X) = \int x f(x) dx = \mu, \quad V(X) = E[X - E(X)]^2 = \int (x - \mu)^2 f(x) dx = \sigma^2$$

새로운 확률변수 Y가 확률변수 X의 선형변환으로 표시된다면 (a와 b는 실수)

$$Y = aX + b$$

그 기대값(평균)과 분산은 다음과 같이 계산된다.

$$E(Y) = E(aX + b)$$

$$= \int (ax + b)f(x)dx$$

$$= a \int xf(x)dx + b$$

$$= aE(X) + b$$

$$= a\mu + b$$

$$V(Y) = Var(aX + b)$$

$$= E[aX + b - E(aX + b)]^2$$

$$= E[a(X - \mu)]^2$$

$$= a^2E(X - \mu)^2$$

$$= a^2\sigma^2$$

E.2. 확률벡터와 분포

확률벡터 \pmb{X} 가 p 차원의 다변량분포를 따른다고 하고 결합확률밀도함수 $f(\pmb{x})=f(x_1,x_2,\dots,x_p)$ 를 를 가진다고 하자.

$$\mathbf{\textit{X}} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ .. \\ X_p \end{bmatrix}$$

다변량 확률벡터의 기대값(평균벡터)과 공분산(행렬)은 다음과 같이 계산된다.

$$\boldsymbol{E}(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

$$V(\pmb{X}) = Cov(\pmb{X}) = E(\pmb{X} - \pmb{\mu})(\pmb{X} - \pmb{\mu})^t = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2p} \\ & \dots & \dots & \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{pp} \end{bmatrix} = \pmb{\Sigma}$$

여기서 $\sigma_{ii}=V(X_i)$, $\sigma_{ij}=Cov(X_i,X_j)=Cov(X_j,X_i)$ 이다. 따라서 공분산 행렬 Σ 는 대칭행렬(symmetric matrix)이다. 다음 공식은 유용한 공식이다.

$$\boldsymbol{\Sigma} = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t = E(\boldsymbol{X}\boldsymbol{X}^t) - \boldsymbol{\mu}\boldsymbol{\mu}^t$$

두 확률변수의 상관계수 ho_{ij} 는 다음과 같이 정의된다.

$$\rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

새로운 확률벡터 Y가 확률벡터 X 의 선형변환라고 하자.

$$Y = AX + b$$

단 여기서 $\pmb{A} = \{a_{ij}\}$ 는 $p \times p$ 실수 행렬이고 $\pmb{b} = (b_1b_2 \dots b_p)^t$ 는 $p \times 1$ 실수 벡터이다.

확률벡터 Y의 기대값(평균벡터)과 공분산은 다음과 같이 계산된다.

$$\begin{split} E(\boldsymbol{Y}) &= E(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) \\ &= \boldsymbol{A}E(\boldsymbol{X}) + \boldsymbol{b} \\ &= \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} \\ V(\boldsymbol{Y}) &= Var(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) \\ &= E[\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} - E(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b})][\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} - E(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b})]^t \\ &= E[\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\mu}][\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\mu}]^t \\ &= E[\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})][\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})]^t \\ &= \boldsymbol{A}E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t]\boldsymbol{A}^t \\ &= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t \end{split}$$

만약 표본 X_i, X_2, \ldots, X_n 이 독립적으로 평균이 μ 이고 공분산이 Σ 인 분포에서 추출되었다면 표본의 평균벡터 \bar{X} 는 평균이 μ 이고 공분산이 $\frac{1}{n}\Sigma$ 인 분포를 따른다.

$$\bar{\mathbf{X}} = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}/n \\ \sum_{i=1}^{n} X_{i2}/n \\ \sum_{i=1}^{n} X_{i3}/n \\ \dots \\ \sum_{i=1}^{n} X_{ip}/n \end{bmatrix}$$

여기서 X_{ij} 는 i 번째 표본벡터 $oldsymbol{X}_i = (X_{i1}X_{i2} \dots X_{ip})^t$ 의 j 번째 확률변수이다.

E.3. 다변량 정규분포

일변량 확률변수 X가 평균이 μ 이고 분산이 σ^2 인 정규분포를 따른다면 다음과 같이 나타내고

$$X \sim N(\mu, \sigma^2)$$

확률밀도함수 f(x) 는 다음과 같이 주어진다.

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2}\right)$$

p-차원 확률벡터 X가 평균이 μ 이고 공분산이 Σ 인 다변량 정규분포를 따른다면 다음과 같이 나타내고

$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

확률밀도함수 $f(\mathbf{x})$ 는 다음과 같이 주어진다.

$$f(\boldsymbol{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^t}{2}\right)$$

다변량 정규분포 $N(\mu, \Sigma)$ 를 따르는 확률벡터 X를 다음과 같이 두 부분으로 나누면

$$m{X} = egin{bmatrix} m{X}_1 \ m{X}_2 \end{bmatrix}, \quad m{X}_1 = egin{bmatrix} m{X}_{11} \ m{X}_{12} \ dots \ m{X}_{1p} \end{bmatrix}, \quad m{X}_2 = egin{bmatrix} m{X}_{21} \ m{X}_{22} \ dots \ m{X}_{2q} \end{bmatrix}$$

각각 다변량 정규분포를 따르고 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} E(\boldsymbol{X}_1) \\ E(\boldsymbol{X}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \begin{bmatrix} V(\boldsymbol{X}_1) & Cov(\boldsymbol{X}_1, \boldsymbol{X}_2) \\ Cov(\boldsymbol{X}_2\boldsymbol{X}_1) & V(\boldsymbol{X}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^t & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim N_{p+q} \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^t & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

확률벡터 $\pmb{X}_2 = \pmb{x}_2$ 가 주어진 경우 \pmb{X}_1 의 조건부 분포는 p-차원 다변량 정규분포를 따르고 평균과 공분산은 다음 과 같다.

$$E(\pmb{X}_1|\pmb{X}_2 = \pmb{x}_2) = \pmb{\mu}_1 + \pmb{\Sigma}_{12}\pmb{\Sigma}_{22}^{-1}(\pmb{\mu}_2 - \pmb{x}_2), \quad V(\pmb{X}_1|\pmb{X}_2 = \pmb{x}_2) = \pmb{\Sigma}_{11} - \pmb{\Sigma}_{12}\pmb{\Sigma}_{22}^{-1}\pmb{\Sigma}_{12}^t$$

예를 들어 2-차원 확률벡터 $\pmb{X}=(X_1,X_2)^t$ 가 평균이 $\pmb{\mu}=(\mu_1,\mu_2)^t$ 이고 공분산 $\pmb{\Sigma}$ 가 다음과 같이 주어진

$$oldsymbol{\Sigma} = egin{bmatrix} \sigma_{11} & \sigma_{12} \ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

이변량 정규분포를 따른다면 확률밀도함수 $f(\boldsymbol{x})$ 에서 \exp 함수의 인자는 다음과 같이 주어진다.

$$\begin{split} &(\pmb{x} - \pmb{\mu}) \pmb{\Sigma}^{-1} (\pmb{x} - \pmb{\mu})^t = \\ &- \frac{1}{2(1-\rho^2)} \left[\left(\frac{(x_1 - \mu_1)^2}{\sigma_{11}} \right) + \left(\frac{(x_2 - \mu_2)^2}{\sigma_{22}} \right) - 2\rho \left(\frac{(x_1 - \mu_1)}{\sqrt{\sigma_{11}}} \right) \left(\frac{(x_2 - \mu_2)}{\sqrt{\sigma_{22}}} \right) \right] \end{split}$$

그리고 p=2인 경우 확률밀도함수의 상수부분은 다음과 같이 주어진다.

$$(2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2} = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}}$$

여기서 $ho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$

만약 $X_2 = x_2$ 가 주어졌을 때 X_1 의 조건부 분포는 정규분포이고 평균과 분산은 다음과 같이 주어진다.

$$E(X_1|X_2=x_2)=\mu_1+\frac{\sigma_{12}}{\sigma_{22}}(\mu_2-x_2)=\mu_1+\rho\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}(\mu_2-x_2)$$

$$V(X_1|X_2=x_2)=\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}=\sigma_{11}(1-\rho^2)$$

다변량 정규분포에서 공분산이 0인 두 확률 변수는 독립이다.

$$\sigma_{ij} = 0 \leftrightarrow X_i$$
 and X_j are independent

E.4. 표준정규분포로의 변환

일변량 확률변수 X가 평균이 μ 이고 분산이 σ^2 인 경우 다음과 같은 선형변환을 고려하면.

$$Z = \frac{X-\mu}{\sigma} = (\sigma^2)^{-1/2}(X-\mu)$$

확률변수 Z 는 평균이 0 이고 분산이 1인 분포를 따른다.

p차원 확률벡터 X 가 평균이 μ 이고 공분산이 Σ 인 분포를 가진다고 가정하자. 공분산 행렬 Σ 는 양정치 행렬 (positive definite matrix)이며 다음과 같은 행렬의 분해가 가능하다.

$$\Sigma = CC^t$$

여기서 C 는 정칙행렬이며 역행렬 C^{-1} 가 존재한다. 위와 같은 행렬의 분해는 스펙트럴 분해(spectral decomposition)을 이용하여 구할 수 있다. 공분산 행렬 Σ 는 양정치 행렬이므로 고유치(eigen value) $(\lambda_1,\lambda_2,\dots,\lambda_p)$ 가 모두 양수이고 정규직교 고유벡터(orthonormal eigen vector)의 행렬 P을 이용하여 다음과 같은 분해가 가능하다.

$$\Sigma = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^t = \mathbf{P} \mathbf{\Lambda}^{1/2} \mathbf{\Lambda}^{1/2} \mathbf{P}^t$$

여기서 Λ 는 고유치 $(\lambda_1,\lambda_2,\dots,\lambda_p)$ 를 대각원소로 가지는 대각행렬이며 $\Lambda^{1/2}$ 는 고유치의 제곱근을 대각원소로 가지는 대각행렬이다. 따라서 $C=P\Lambda^{1/2}$ 로 하면 위와 같은 행렬의 분해가 가능하다. 정규직교 고유벡터(orthonormal eigen vector)의 행렬 P는 직교행렬이므로

$$C^{-1} = (P\Lambda^{1/2})^{-1} = \Lambda^{-1/2}P^t$$

p차원 확률벡터 \boldsymbol{X} 의 다음과 같은 선형변환을 고려하면.

$$Z = C^{-1}(X - \mu) = \Lambda^{-1/2}P^{t}(X - \mu)$$

확률벡터 Z 는 평균이 0 이고 공분산이 I 인 분포를 따른다 (why?).

확률벡터 X가 정규분포를 따른다면 선형변화한 확률벡터 Z도 정규분포를 따른다.

E.5. 예제

예를 들어 이변량확률벡터 X가 다음과 같은 평균벡터와 공분산을 가진 정규분포를 따른다고 하자

$$\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

공분산행렬 σ 의 고유치는 $|\sigma - \lambda I| = 0$ 의 방정식을 풀어 구할 수 있다.

$$|\sigma - \lambda I| = \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix} = \lambda^2 - 4\lambda + 3 = 0$$

방정식을 풀면 고유치는 $(\lambda_1, \lambda_2) = (3, 1)$ 이다. 각 고유치에 대한 고유벡터 $\mathbf{p} = (p_1, p_2)^t$ 는 $\mathbf{\Sigma} \mathbf{p} = \lambda \mathbf{p}$ 으로 구할 수 있다. 각 고유치에 대하여 방정식을 구하면 다음 두 개의 방정식을 얻을 수 있다.

$$p_1 - p_2 = 1$$
 and $p_1 + p_2 = 0$

정규직교 벡터의 조건을 만족 시키기 위해서 $p_1^2+p_2^2=1$ 의 조건을 적용하면 다음과 같은 정규직교 고유행렬을 얻을 수 있다.

$$m{P} = egin{bmatrix} rac{1}{\sqrt{2}} & -rac{1}{\sqrt{2}} \ rac{1}{\sqrt{2}} & rac{1}{\sqrt{2}} \end{bmatrix}$$

또한

$$\mathbf{\Lambda} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{\Lambda}^{1/2} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

따라서 $C^{-1} = \Lambda^{-1/2} \boldsymbol{P}^t$ 이며

$$m{C}^{-1} = m{\Lambda}^{-1/2} m{P}^t = egin{bmatrix} rac{1}{\sqrt{3}} & 0 \ 0 & 1 \end{bmatrix} egin{bmatrix} rac{1}{\sqrt{2}} & rac{1}{\sqrt{2}} \ -rac{1}{\sqrt{2}} & rac{1}{\sqrt{2}} \end{bmatrix} = egin{bmatrix} rac{1}{\sqrt{6}} & rac{1}{\sqrt{6}} \ -rac{1}{\sqrt{2}} & rac{1}{\sqrt{2}} \end{bmatrix}$$

F. 이차형식과 제곱합의 분포

이 장에서는 선형 회귀모형의 추론과 검정에 자주 사용되는 제곱합의 분포와 관련이 있는 이차형식의 분포의 성질을 살펴본다.

F.1. 이차형식

n-차원 벡터 $\pmb{x}^t = [x_1, x_2, \dots, x_n]$ 과 대칭행렬 \pmb{A} 에 대하여 이차형식(quadratic form)은 다음과 같이 정의된다.

$$Q_A(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \tag{F.1}$$

이차형식의 정의에서 반드시 행렬 A를 대칭행렬로 정의하지 않아도 되지만 임의의 행렬에 대하여 이차형식의 값이 동일한 대칭행렬이 존재하기 때문에 정의에서 이차형식으로 국한하는 것이 일반적이다.

정의 $\mathbf{F.1}$ (양정치 행렬). 이차형식 $Q_A(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ 가 영벡터가 아닌 모든 벡터 \mathbf{x} 에 대하여 0 보다 크면, 즉

$$\boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} > 0$$
 for all $\boldsymbol{x} \in \mathbb{R}^n$

A를 양정치(positive definite)라고 부른다.

만약 이차형식 $Q_A({\pmb x}) = {\pmb x}^t {\pmb A} {\pmb x}$ 가 영벡터가 아닌 모든 벡터 ${\pmb x}$ 에 대하여 0 보다 크거나 같다면

$$\boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} > 0$$
 for all $\boldsymbol{x} \in \mathbb{R}^n$

A를 양반정치(positive semi-definite)라고 부른다.

정칙행렬 B에 대하여 다음과 같은 선형변화을 고려하자.

$$\boldsymbol{x} = \boldsymbol{B} \boldsymbol{y}$$
 or $\boldsymbol{y} = \boldsymbol{B}^{-1} \boldsymbol{x}$

벡터 x로 정의된 이차형식은 벡터 u의 형태로 다음과 같이 변화할 수 있다.

$$Q(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} = \mathbf{y}^t \mathbf{B}^t \mathbf{A} \mathbf{B} \mathbf{y} = Q^*(\mathbf{y})$$

이차형식의 성질은 정칙 선형변환에서 유지된다. 즉 행렬 $m{A}$ 가 양(반)정치 행렬이고 행렬 $m{B}$ 가 정칙행렬이면 행렬 $m{B}^t m{A} m{B}$ 도 양(반)정치 행렬이다.

F.2. 대칭행렬의 대각화

n차원 대칭행렬 A 에 대하여 직교행렬 P가 존재하여 다음과 같은 분해가 가능하다.

$$\mathbf{P}^{t}\mathbf{AP} = \mathbf{\Lambda} = diag(\lambda_{1}, \lambda_{2}, \dots, \lambda_{n}) \tag{F.2}$$

식 $\mathbf{F}.2$ 의 분해에서 λ_i 는 행렬 $m{A}$ 의 고유치이며 행렬 $m{P}$ 의 i 번째 열은 대응하는 고유벡터 $m{p}_i$ 로 구성되어 있다.

$$P = [\boldsymbol{p}_1 \ \boldsymbol{p}_2 \ \dots \ \boldsymbol{p}_n]$$

이제 위의 분해를 증명해 보자. 고유치 λ_i 와 대응하는 고유벡터 \pmb{p}_i 의 정의에 따라서 다음과 같은 n개의 식을 얻을 수 있고

$$Ap_i = \lambda_i p_i, \quad i = 1, 2, 3 \dots, n$$

위의 식을을 합쳐서 표기하면 다음과 같은 식을 얻으며 이는 식 F.2 를 의미한다.

$$AP = P\Lambda$$

식 F.2 를 다시 쓰면 다음과 같은 스펙트럴 분해(spectral decomposition)를 얻는다.

$$\boldsymbol{A} = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^t = \sum_{i=1}^n \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i^t$$
 (F.3)

참고로 대각합과 행렬식에 대한 고유치위 관계를 나타내는 다음의 유용한 두 식을 반드시 기억하자.

$$tr(\pmb{A}) = \sum_i \lambda_i, \quad |\pmb{A}| = \prod_i \lambda_i$$

대칭행렬의 분해 식 F.2 를 이용하면 다음과 같은 이차형식의 분해를 얻을 수 있다.

$$Q(\boldsymbol{x}) = \boldsymbol{x}^{t} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^{t} \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^{t} \boldsymbol{x} = \boldsymbol{y}^{t} \boldsymbol{\Lambda} \boldsymbol{y} = \sum_{i=1}^{n} \lambda_{i} y_{i}^{2}$$
 (F.4)

위의 식에서

$$oldsymbol{y} = oldsymbol{P}^t oldsymbol{x}$$

이차형식의 분해식 식 F.4 를 보면 행렬 \boldsymbol{A} 의 모든 고유치가 0보다 크면 양정치 임을 알 수 있다. 또한 모든 고유치가 0보다 크거나 같으면 양반정치 임을 알 수 있다.

또한 $rank(\mathbf{A}) = rank(\mathbf{\Lambda})$ 이며 이는 0이 아닌 고유치의 개수가 행렬 \mathbf{A} 의 계수(rank)임을 알 수 있다.

F.3. 멱등행렬

n-차원 행렬 A 가 다음과 같은 성질을 가지면 멱등행렬(idenpotent matrix)라고 부른다.

$$A^2 = AA = A$$

멱등행렬은 다음과 같은 성질을 가지고 있다.

- 멱등행렬의 고유치는 0 또는 1이다.
- 멱등행렬은 대각합이 계수와 같다.

$$tr(\mathbf{A}) = rank(\mathbf{A})$$

- 멱등행렬은 양반정치 행렬이다.
- A 멱등행렬이면 I A도 멱등행렬이다.

특별히 대칭인 멱등행렬을 사영행렬(또는 투영행렬, projection matrix)라고 부른다.

최소제곱법에서 식 B.10 에서 나타난 행렬 $\boldsymbol{H} = \boldsymbol{X} (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t$ 는 멱등행렬이며 따라서 사영행렬이다.

F.4. 이차형식의 분포

F.4.1. 카이제곱 분포

만약 확률변수 x 이 표준 정규분포 N(0,1) 을 따른다면 $y=x^2$ 은 자유도가 1인 카이제곱 분포 χ_1^2 를 따른다. 더 나아가 n 개의 확률변수 x_1,x_2,\dots,x_n 이 서로 독립이고 표준 정규분포 N(0,1) 을 따른다면 제곱합 $v=x_1^2+x_2^2+\dots+x_n^2$ 은 자유도가 n인 카이제곱 분포 χ_n^2 를 따른다.

이렇게 카이제곱 분포는 표준 정규분포를 따르는 서로 독립인 확률변수들의 제곱값에 대한 분포이다.

F.4.2. 비중심 카이제곱 분포

만약 확률변수 x가 $N(\mu,1)$ 을 따른다면 $v=x^2$ 은 자유도가 1인 비중심 카이제곱 분포, $\chi_1^2(\lambda^2)$ 를 따른다. 여기서 비중심 카이제곱 분포의 자유도는 1이고 비중심모수 $\lambda^2=\mu^2$ 으로 주어진다.

이제 n개의 서로 독립인 확률 변수 x_1,x_2,\cdots,x_n 이 각각 $N(\mu_i,1)$ 을 따른다면 $v=x_1^2+\cdots+x_n^2$ 은 자유도가 n이고 비중심 모수가 $\lambda^2=\sum_{i=1}^n\mu_i^2$ 인 비중심 카이제곱 분포, $\chi_n^2(\lambda^2)$ 를 따른다.

참고로 확률변수 x가 N(0,1)을 따른다면 $v=x^2$ 은 중심 카이제곱 분포, χ_1^2 를 따르며 이때는 비중심모수가 $\lambda^2=0$ 이다. 즉, 비중심모수가 0인 비중심 카이제곱 분포(non-central chi square distribution)를 중심 카이제곱 분포(central chi square distribution)라고 한다. 또한 중심 카이제곱 분포는 중심을 빼고 카이제곱 분포라고 부른다.

F.4.3. 이차형식의 분포

n개의 서로 독립인 확률 변수 x_1,x_2,\cdots,x_n 이 각각 $N(\mu_i,\sigma^2)$ 를 따른다면 n-차원의 확률벡터 ${\pmb x}$ 는 다음과 같은 다변량 정규분포를 따른다고 할 수 있다.

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$$

위에서 $\pmb{\mu}^t = (\mu_1, \mu_2, \dots, \mu_n)$

이제 이차형식의 분포에 대하여 논의하자.

정리 $\mathbf{F.1}$ (이차형식의 분포). n-차원의 확률벡터 \mathbf{x} 가 $N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ 를 따른다면 이차형식 $Q = \mathbf{x}^t \mathbf{A} \mathbf{x}$ 의 분포는 다음 과 같다.

$$V = \frac{Q}{\sigma^2} = \frac{\boldsymbol{x}^t A \boldsymbol{x}}{\sigma^2} \equiv_d \sum_{i=1}^n \lambda_i \frac{u_i^2}{\sigma^2}$$
 (F.5)

위의 식에서 $x \equiv_d y$ 는 확률변수 x와 y가 동일한 분포를 가진다는 것을 의미한다.

식 F.5 에서 행렬 A의 스펙트럴 분해는 $A=P\Lambda P^t$ 이며 λ_i 는 행렬 A의 고유치, 즉 행렬 Λ 의 대각원소이다. 또한 확률변수 u_i 들은 서로 독립이며 정규분포 $N(\eta_i,\sigma^2)$ 를 따른다. 여기서 $\eta_1,\eta_2,\ldots,\eta_n$ 는 다음과 같이 정의된다.

$$oldsymbol{\eta} = egin{bmatrix} \eta_1 \ \eta_2 \ dots \ \eta_n \end{bmatrix} = oldsymbol{P}^t oldsymbol{\mu}$$

즉, 식 F.5 에서 $u_1^2/\sigma^2, u_2^2/\sigma^2, \dots, u_n^2/\sigma^2$ 는 서로 독립이며 각각 자유도가 1 이고 비중심 모수가 $\eta_1^2/\sigma^2, \eta_2^2/\sigma^2, \dots, \eta_n^2/\sigma^2$ 인 비중심 카이제곱-분포를 따른다.

정리 F.1 의 식 F.5 에서 나타난 이차형식의 분포는 비중심 카이제곱 분포를 따르는 서로 독립인 확률 변수들의 가중 평균과 같다는 것이다.

이는 이차형식의 분포가 비중심 카이제곱 분포를 따른다는 것이 아님을 주의해야 한다. 그러면 어느 경우에 이차형식의 분포가 비중심 카이제곱 분포를 따르는가 생각해 보자.

가장 쉽게 생각할 수 있는 경우가 식 ${
m F.5}$ 에서 λ_i 들의 값들이 0 또는 1인 경우이다. 이러한 경우는 행렬 ${m A}$ 가 멱등행렬인 경우이다. 실제로 다음 정리는 이차형식의 분포가 비중심 카이제곱 분포를 따르는 필요충분 조건 이 행렬

A 가 멱등행렬이라는 것을 말해준다.

따름정리 F.1. n-차원의 확률벡터 \boldsymbol{x} 가 $N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ 를 따른다면 이차형식 $Q = \boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x}$ 의 분포가 자유도가 r 이 며 다음과 같은 비중심 모수 λ^2 을 가지는 비중심 카이제곱 분포를 따르는 필요충분 조건은 \boldsymbol{A} 가 멱등행렬이고 $rank(\boldsymbol{A}) = r$ 인 경우이다.

$$\lambda^2 = \frac{\boldsymbol{\mu}^t \boldsymbol{A} \boldsymbol{\mu}}{\sigma^2}$$

더 나아가 $A\mu = 0$ 이면 이차형식의 분포는 자유도가 r인 (중심)카이제곱 분포를 따른다.

F.4.4. 이차형식의 독립

두 개의 이차형식이 독립일 조건은 다음 정리와 같다.

정리 $\mathbf{F.2}$ (이차형식의 독립). n-차원의 확률벡터 \mathbf{x} 가 $N\left(\boldsymbol{\mu},\sigma^2\mathbf{I}\right)$ 를 따른다고 하자. 두 이차형식 $Q_1=\mathbf{x}^t\mathbf{A}\mathbf{x}$ 과 $Q_2=\mathbf{x}^t\mathbf{B}\mathbf{x}$ 가 서로 독립일 필요충분 조건은 $\mathbf{A}\mathbf{B}=\mathbf{0}$ 이다.

F.4.5. 이차형식의 차이

만약 3 개의 이차형식 Q, Q_1, Q_2 가 있어서 다음과 같은 관계가 있다고 하자.

$$Q = \mathbf{x}^t \mathbf{A} \mathbf{x} = Q_1 + Q_2 = \mathbf{x}^t \mathbf{A}_1 \mathbf{x} + \mathbf{x}^t \mathbf{A}_2 \mathbf{x}$$

이러한 경우 두 이차형식 Q 과 Q_1 이 각각 카이제곱 분포를 따를 때 $Q_2=Q-Q_1$ 이 카이제곱 분포를 따르는 조건이 중요하다. 다음 정리는 그 조건을 행렬 ${m A}_2$ 가 양반정치인 경우라는 것을 말해준다.

정리 F.3 (이차형식의 차이). n-차원의 확률벡터 x 가 $N\left(\mu,\sigma^2I\right)$ 를 따른다고 하자. 세 개의 이차형식 $Q=x^tAx,Q_1=x^tA_1x,Q_2=x^tA_2x$ 가 있다고 하고 $Q=Q_1+Q_2$ 인 관계를 가진다고 가정하자.

만약 Q/σ^2 이 $\chi^2_r(\lambda^2)$ 을 따르고 Q_1/σ^2 이 $\chi^2_{r_1}(\lambda^2_1)$ 을 따르며 행렬 \pmb{A}_2 가 양반정치 행렬이면 다음을 만족한다. 두 이차형식 Q_1 과 Q_2 는 서로 독립이다. 또한 이차형식 Q_2 는 자유도가 $r_2=r-r_1$ 이고 비중심 모수가 $\lambda^2_2=\lambda^2-\lambda^2_1$ 인 비중심 카이제곱분포를 따른다.

F.5. 코크란의 정리

선형모형에서 자주 등장하는 제곱합들의 분해, 즉 이차형식의 분해를 생각할 때 각 제곱합들의 분포를 아는 것이 매우 중요하다. 다음에 제시된 코크란의 정리(Cochran's Theorem)는 총 제곱합을 분해했을 때 각 제곱합의 분포가 카이제곱 분포를 따를 조건을 말해준다.

정리 $\mathbf{F.4}$ (COCHRAN'S THEOREM). n-차원의 확률벡터 \mathbf{x} 가 $N\left(\boldsymbol{\mu},\sigma^2\boldsymbol{I}\right)$ 를 따른다고 하자. k 개의 이차형식 $Q_j=\mathbf{x}^t\boldsymbol{A}_j\mathbf{x}, j=1,2,\ldots,k$ 를 생각하고 다음과 같은 관계를 가진다고 하자.

$$oldsymbol{x}^toldsymbol{x} = \sum_{i=1}^n x_i^2 = \sum_{j=1}^k Q_j$$

즉, $\sum_{j=1}^k \pmb{A}_j = \pmb{I}$ 이다. 또한 $r_j = \mathrm{rank}\left(\pmb{A}_j\right)$ 이고 $\lambda_j^2 = \pmb{\mu}^t \pmb{A}_j \pmb{\mu}$ 라고 하자.

k 개의 이차형식 Q_1,Q_2,\dots,Q_k 들이 모두 독립이고 각 이차형식 Q_j/σ^2 가 비중심 카이제곱 분포 $\chi^2_{r_j}\left(\lambda^2_j\right)$ 를 따를 필요충분 조건은 다음과 같다.

$$r_1 + r_2 + \dots + r_k = n$$

이제 제곱합의 분포들에 대하여 지금까지 학습한 내용을 정리해보자. 만약 n-차원의 확률벡터 \boldsymbol{x} 가 $N\left(\boldsymbol{\mu},\sigma^2\boldsymbol{I}\right)$ 를 따른다고 하고 위의 코크란의 정리와 같이 제곱합의 분해를 고려하자. 다음에 제시된 모든 문장은 서로 동치 (equivalent)이다.

- 1. 이차형식 Q_1,Q_2,\ldots,Q_k 들이 모두 독립이다.
- 2. 모든 $j=1,2,\ldots,k$ 에 대하여 이차형식 Q_j/σ^2 가 비중심 카이제곱 분포 $\chi^2_{r_j}\left(\lambda^2_j\right)$ 를 따른다.
- 3. $A_1, A_2, ..., A_k$ 가 모두 멱등행렬이다.
- 4. 모든 $j \neq k$ 에 대하여 $\boldsymbol{A}_{i}\boldsymbol{A}_{k} = \boldsymbol{0}$ 이다.
- 5. $r_1 + r_2 + \dots + r_k = n$