

## 4강. 중회귀모형 (1)

◆ 담당교수 : 김성수 교수

### ■ 주요용어

용어	해설
중회귀모형	<p>독립변수의 수가 <math>k</math>개인 중회귀모형은 일반적으로 다음과 같이 표현된다.</p> $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i$ <p><math>\beta_0, \beta_1, \dots, \beta_k</math> : 회귀계수  <math>\epsilon_i \sim N(0, \sigma^2)</math> 이고 서로 독립  <math>i = 1, 2, \dots, n</math></p> <p>여기서 <math>\beta_0, \beta_1, \dots, \beta_k</math> 는 모집단의 회귀계수이고, <math>\epsilon_i</math> 는 반응변수 <math>Y_i</math>를 측정할 때 발생하는 오차이다.</p>
햇행렬	<p>중회귀모형에서 추정값과 잔차벡터는</p> $\begin{aligned}\hat{Y} &= Xb \\ &= X(X'X)^{-1}X'Y \\ &= HY \\ e &= Y - \hat{Y} = Y - Xb \\ &= Y - X(X'X)^{-1}X'Y \\ &= (I - X(X'X)^{-1}X')Y \\ &= (I - H)Y\end{aligned}$ <p>이 된다. 여기서 <math>H = X(X'X)^{-1}X'</math> 을 나타내며, 이를 햇행렬(hat matrix) 이라 한다.</p>
잔차	<p>중회귀모형을 데이터에 적합시켜 얻은 추정된 회귀식의 값 <math>\hat{Y}_i</math>과 관찰값 <math>Y_i</math>의 차이를 잔차, <math>e_i = Y_i - \hat{Y}_i</math> 라고 하며, 잔차벡터는 다음과 같이 간단히 표시된다.</p> $\begin{aligned}e &= Y - \hat{Y} = Y - Xb \\ &= Y - X(X'X)^{-1}X'Y \\ &= (I - X(X'X)^{-1}X')Y \\ &= (I - H)Y\end{aligned}$
총제곱합	<p>총변동은 총제곱합 (total sum of squares, <math>SST</math>)으로,</p> $SST = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n(\bar{Y})^2 = Y'Y - n(\bar{Y})^2$ <p>이를 행렬로 표현하면 다음과 같다.</p> $\begin{aligned}SST &= Y'Y - n(\bar{Y})^2 \\ &= Y'(I - \frac{J}{n})Y\end{aligned}$ <p>여기서 <math>J</math> 행렬은 모든 요소가 1인 <math>n \times n</math> 행렬을 의미한다. <math>SST</math>에 대한 자유도는 <math>n-1</math>이다.</p>
잔차제곱합	<p>잔차제곱합(sum of squares due to residual errors, <math>SSE</math>)은 다음과 같다.</p> $\begin{aligned}SSE &= \sum (Y_i - \hat{Y}_i)^2 = \sum e_i^2 = e'e \\ &= [(I - H)Y]'[(I - H)Y] \\ &= Y'(I - H)'(I - H)Y \\ &= Y'(I - H)Y\end{aligned}$ <p><math>SSE</math>에 대한 자유도는 <math>n-k-1</math>이다.</p>

회귀제곱합	<p>회귀제곱합 (sum of squares due to regression, <math>SSR</math>)은</p> $  \begin{aligned}  SSR &= \sum (\hat{Y}_i - \bar{Y})^2 \\  &= \sum \hat{Y}_i^2 - n(\bar{Y})^2 \\  &= \hat{Y}'\hat{Y} - n(\bar{Y})^2  \end{aligned}  $ <p>이고, <math>\hat{Y} = Xb</math>로부터</p> $  \begin{aligned}  SSR &= b'X'Y - n(\bar{Y})^2 \\  &= Y'HY - Y'\left(\frac{1}{n}\right)Y \\  &= Y'\left(H - \frac{1}{n}\right)Y  \end{aligned}  $ <p>이다. <math>SSR</math>의 자유도는 독립변수의 수와 같은 <math>k</math>가 된다.</p>
-------	--

### ■ 연습문제

1. 중회귀모형  $Y = X\beta + \epsilon$ 의 최소제곱추정치  $b$ 를  $Y$ 와  $X$ 의 행렬식으로 표현하면?

- ①  $b = (X'X)^{-1}XY$
- ②  $b = (X'X)^{-1}X'Y$
- ③  $b = X(X'X)^{-1}XY$
- ④  $b = X(X'X)^{-1}X'Y$

정답 및 해설 : ②

2. 다음 중 잔차제곱합 SSE를 나타내는 것은? 단,  $H = X(X'X)^{-1}X'$

- ①  $Y'(I - H)Y$
- ②  $Y'(H - I)Y$
- ③  $Y'HY$
- ④  $Y'Y - H$

정답 및 해설 : ①

3. 다음 중  $Var(Y) = I\sigma^2$  일 때, 잔차  $e$ 의 분산 - 공분산 행렬은?

- ①  $(I - H)X\sigma^2$
- ②  $I\sigma^2$
- ③  $(I - H)\sigma^2$
- ④  $(H - I)\sigma^2$

정답 및 해설 : ③

4. 중회귀모형  $Y = X\beta + \epsilon$  에서  $Var(\epsilon) = I\sigma^2$  일 때,  $\beta$ 의 최소제곱추정치  $b$ 의 분산 - 공분산 행렬은?

- ①  $X(X'X)^{-1}\sigma^2$
- ②  $(X'X)^{-1}\sigma^2$
- ③  $X(X'X)^{-1}X'\sigma^2$
- ④  $X'X\sigma^2$

정답 및 해설 : ②

#### ■ 참고사이트

- 강명욱, 김영일, 안철환, 이용구, 『회귀분석』, 을곡출판사, 1996.
- 박성현, 『회귀분석』 (제3판), 민영사, 2007.
- Faraway, J.J. (2002), Practical Regression and Anova Using R, (www.google.com에서 검색 후, pdf 파일로 다운받을 수 있음)
- Peter Dalgaard (2005), Introductory Statistics with R, Springer, (www.google.com에서 검색 후, pdf 파일로 다운받을 수 있음)
- R 사이트 바로가기

<https://www.r-project.org/>

- R Studio 사이트 바로가기

<https://www.rstudio.com/>