회귀분석

선형관계의 강도

- 관측된 종속변수값(관측값): y_i
- 선형관계로 설명되는 y 부분(예측값): $\hat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x$
- 선형관계로 설명되지 않는 y 부분(잔차): $e_i = y_i - \hat{y_i} = y_i - (\widehat{\beta_0} + \widehat{\beta_1}x)$
- $SSE = \sum_{i=1}^{n} e_i{}^2 = \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_i)^2 = S_{yy} - \frac{S_{xy}{}^2}{S_{xx}}$

선형관계의 강도

- $SSE = \sum_{i=1}^{n} e_i^{\ 2} = \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i)^2 = S_{yy} - \frac{S_{xy}{}^2}{S_{xx}}$

- $SST = S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$

- $S_{yy} = \left(S_{yy} - \frac{S_{xy}{}^2}{S_{xx}}\right) + \frac{S_{xy}{}^2}{S_{xx}} = SSE + \frac{S_{xy}{}^2}{S_{xx}}$

- $\frac{S_{xy}{}^2}{S_{xx}}$: 자료의 y 값들의 총변동 중에서 회귀모형에 의해서 설명될 수 있는 부분(회귀제곱합: SSR)

선형관계의 강도

- 총제곱합의 분해:

$$S_{yy}(\text{y의 총변동}: SST) = \frac{{S_{xy}}^2}{S_{xx}}(\text{선형관계로 설명 되는 변동}: SSR) + \left(S_{yy} - \frac{{S_{xy}}^2}{S_{xx}}\right)(\text{선형관계로 설명되지 않는 변동: SSE})$$

- 결정계수:

$$R^2 = \frac{SSR}{SST} = \frac{{S_{xy}}^2}{S_{xx}S_{yy}} = 1 - \frac{SSE}{SST}$$

- 표본상관계수: $\mathrm{r} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad \Rightarrow \quad R^2 = r^2$

예제

- 예제. 다음 표는 어떤 알레르기 증세에 효과가 있다고 하는 새로 개발된 약품의 복용량(mg)과 효과가 지속되는 기간(일)을 기록한 자료이다.

복용량(x)	3	3	4	5	6	6	7	8	8	9
지속기간(y)	9	5	12	9	14	16	22	18	24	22

잔차의 검토

- 단순선형회귀모형

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

오차 ε_i에 대한 가정

(1) ε_i의 평균은 0이다.

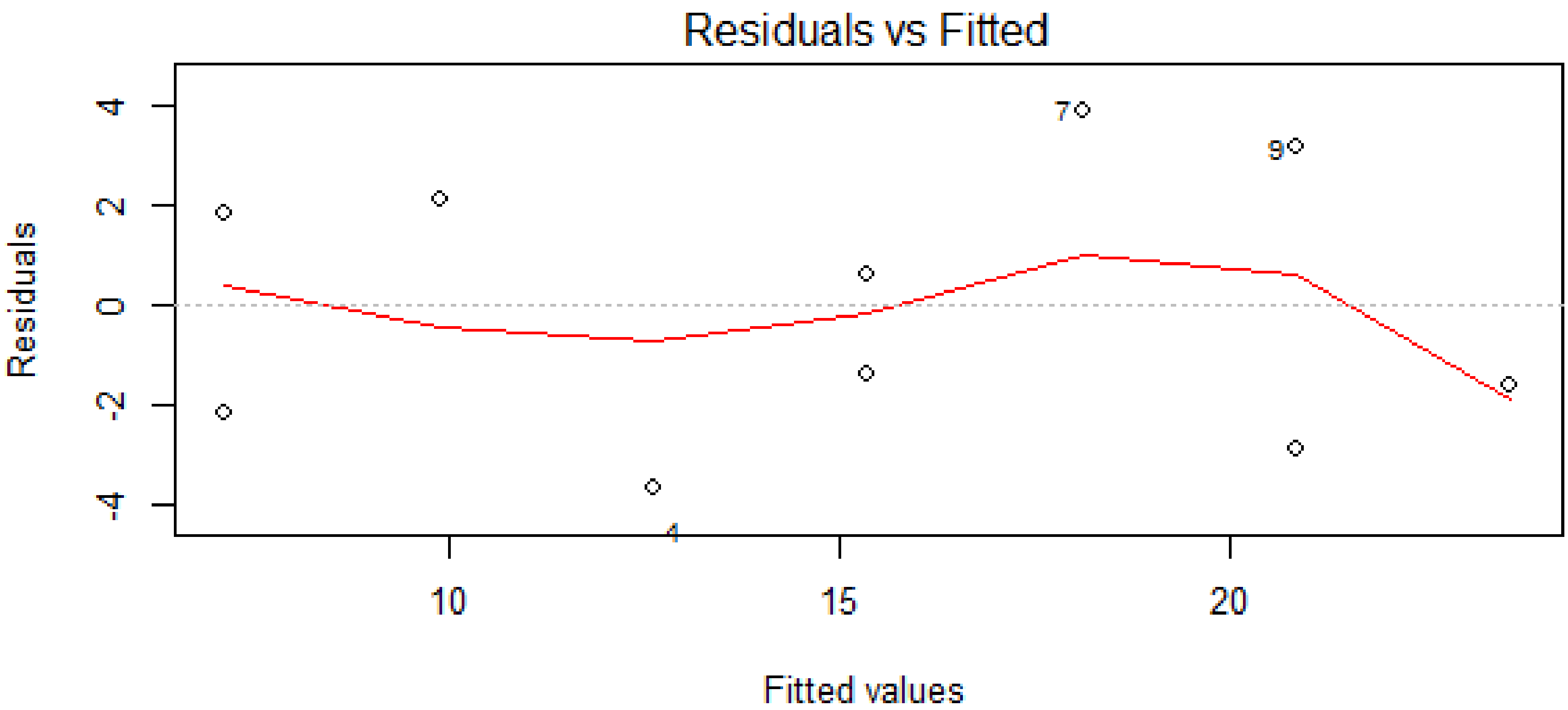
(2) ε_i들은 서로 독립이다. (독립성)

(3) ε_i의 분산은 σ^2이다. (등분산성)

(4) ε_i는 정규분포 $N(0, \sigma^2)$을 따른다. (정규성)

잔차의 검토 – 선형성, 등분산성 검정

➢ 잔차 vs. 적합값(예측값)



잔차의 검토 – 선형성, 등분산성 검정

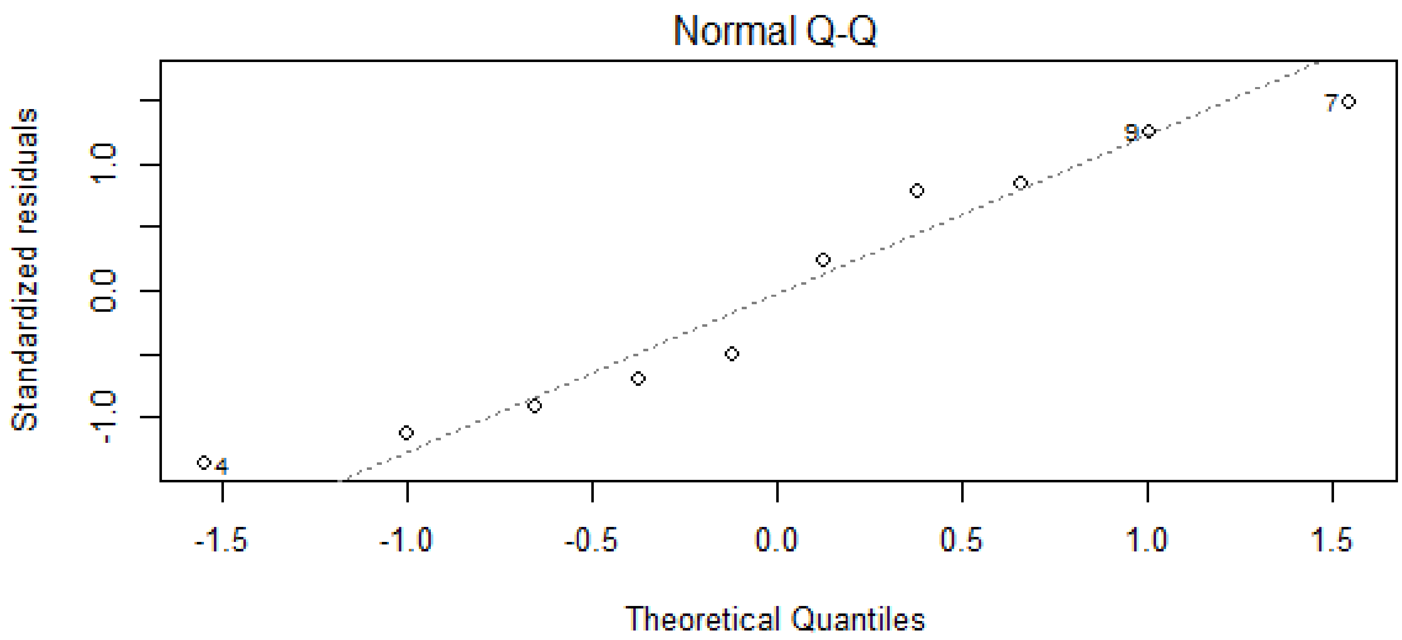
- 잔차 vs. 독립변수값

$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$: 선형 관계

잔차의 검토 – 오차의 정규성 검정

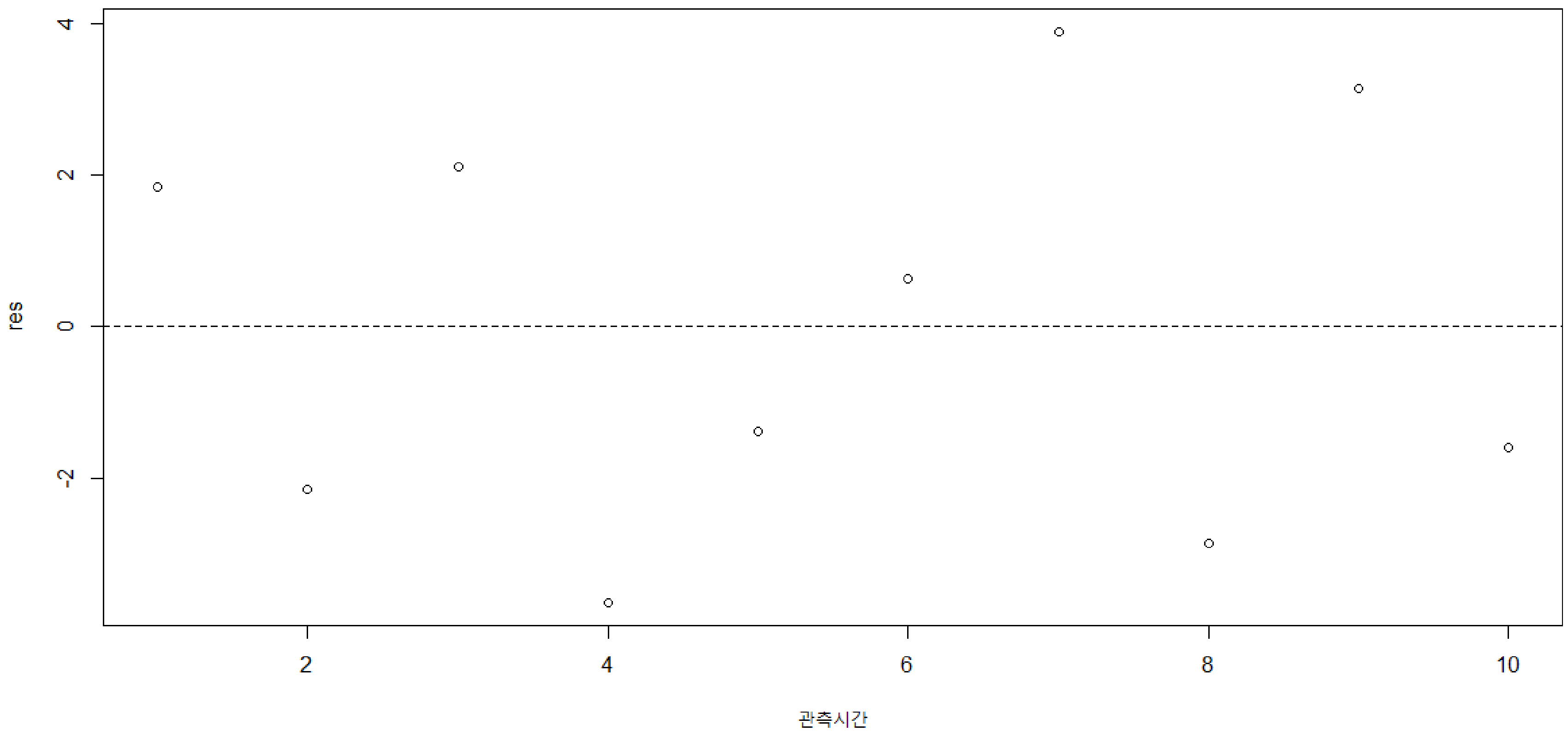
- 잔차의 히스토그램

- 잔차의 정규확률그림



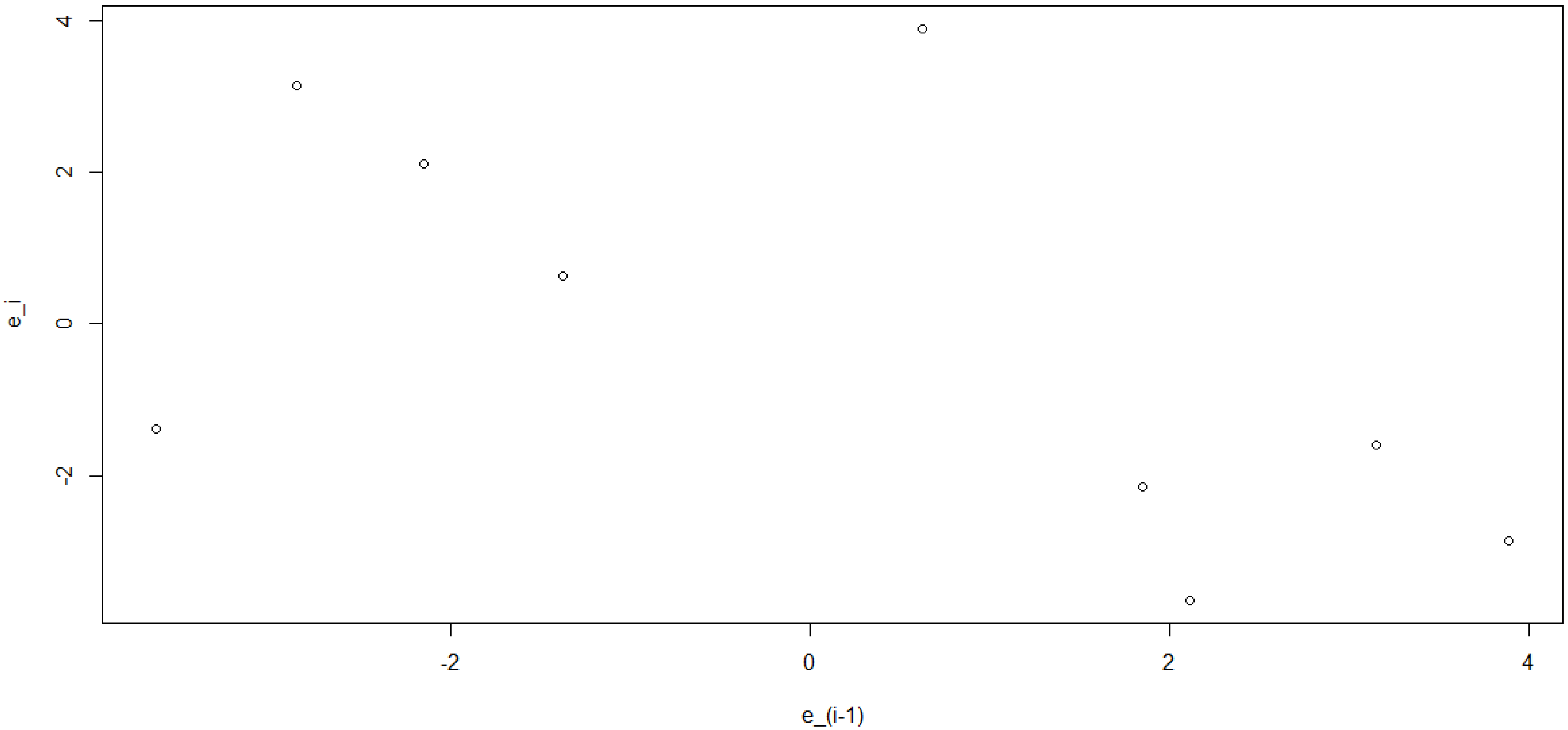
잔차의 검토 – 오차의 독립성 검정

- 관측시간에 따른 잔차의 산점도



잔차의 검토 – 오차의 독립성 검정

- e_i와 e_{i-1}의 산점도



수고하셨습니다.

- 과제 X