

10강. 모형진단

◆ 담당교수 : 김성수 교수

■ 주요용어

용어	해설
오차의 등분산성	회귀모형에서 모든 관측값에 대해 오차의 분산이 같다. 즉, $Var(\epsilon_i) = \sigma^2$ 을 말한다.
회귀모형의 선형성	회귀모형에서 반응변수와 일부 설명변수 사이의 관계가 선형이 되는 것을 말한다.
오차의 정규성	회귀모형에서 오차가 정규분포를 따른다는 가정을 말한다.

■ 연습문제

1. 오차의 등분산성을 검정하고자 한다. R 함수 (a)는 ?

```
> install.packages("car")
> library(car)
> ( a )(goose.lm)
Non-constant Variance Score Test
Variance formula : ~ fitted.values
Chisquare = 81.41318 Df = 1 p = 1.831324e-19
```

정답 및 해설 : ncvTest

2. 오차의 정규성을 검정하고자 한다. 정규확률그림을 그리기 위한 함수 (b)는 ?

```
> library(car)
> goose.lm = lm(photo ~ obsA, data=goose)
> ( b )(goose.lm)
```

정답 및 해설 : qqPlot

3. Shapiro-Wilk의 정규성 검정을 실시하였다. 다음 결과에서 정규성검정 결과는?
(단, 유의수준 0.05 기준)

```
> install.packages("mvnrmtest")
> library(mvnrmtest)
> goose.rstudent = rstudent(goose.lm)
> shapiro.test(goose.rstudent)
Shapiro-Wilk normality test
data:  goose.rstudent
W = 0.7192, p-value = 5.971e-08
```

정답 및 해설 : $p\text{-값} = 5.97 \times 10^{-8} < 0.05$ 이므로 정규분포를 따른다는 가설을 기각한다.

■ 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 <http://www.r-project.org/>
- R Studio 사이트 <https://www.rstudio.com/>