Daeyoung Lim
Grad School
Year of 2010, late

1. 다음과 같은 다중선형회귀모형

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

을 고려하자. 여기서, Y는 독립변수를, $X_1, \ldots, X_p$는 설명변수들을, 그리고 $\beta_0, \beta_1, \ldots, \beta_p$는 회귀계수들을 의미하며 ϵ은 평균이 0, 분산이 σ^2인 오차항을 의미한다.

(a) 이 선형회귀모형을 적합한 후 얻어진 예측값을 $\widehat{Y}$으로 나타낼 때, $\widehat{Y}$의 기댓값 및 분산을 구하시오.

(b) 이 회귀모형의 잔차(residual)를 $e = Y - \widehat{Y}$으로 나타낼 때, e의 기댓값 및 분산을 구하시오.

(c) 잔차제곱합(residual sum of squares)의 기댓값을 구하시오.

Solution:

(a) 행렬꼴로 바꾸면 쉽다.

$$\begin{aligned}
\mathrm{E}\left(\widehat{Y}\right) &= \mathrm{E}\left(\mathbf{X}\widehat{\beta}\right) && (1)\\
&= \mathbf{X}\beta && (2)\\
\mathrm{Var}\left(\widehat{Y}\right) &= \mathrm{Var}\left(\mathbf{X}\widehat{\beta}\right) && (3)\\
&= \mathbf{X}\mathrm{Var}\left(\widehat{\beta}\right)\mathbf{X}' && (4)\\
&= \sigma^2 \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' && (5)
\end{aligned}$$

(b) 이미 구해놨으니 더 쉽다.

$$\begin{aligned}
\mathrm{E}\left(Y - \widehat{Y}\right) &= \mathbf{X}\beta - \mathbf{X}\beta && (6)\\
&= 0 && (7)\\
\mathrm{Var}\left(Y - \widehat{Y}\right) &= \mathrm{Var}\left(\left(\mathbf{I} - \mathbf{H}\right)Y\right) && (8)\\
&= \left(\mathbf{I} - \mathbf{H}\right)\sigma^2\mathbf{I}\left(\mathbf{I} - \mathbf{H}\right)' && (9)\\
&= \sigma^2\left(\mathbf{I} - \mathbf{H}\right) && (10)
\end{aligned}$$

(c) 잔차제곱합의 기댓값은 다음과 같다.

$$\begin{aligned}
\left(Y - \mathbf{H}Y\right)'\left(Y - \mathbf{H}Y\right) &= Y'\left(\mathbf{I} - \mathbf{H}\right)Y && (11)\\
&= \left(Y - \mathbf{X}\beta\right)'\left(\mathbf{I} - \mathbf{H}\right)\left(Y - \mathbf{X}\beta\right) && (12)\\
&= \epsilon'\left(\mathbf{I} - \mathbf{H}\right)\epsilon && (13)\\
\mathrm{E}\left(\epsilon'\left(\mathbf{I} - \mathbf{H}\right)\epsilon\right) &= \mathrm{Tr}\left(\left(\mathbf{I} - \mathbf{H}\right)\mathrm{E}\left(\epsilon\epsilon'\right)\right) && (14)\\
&= \sigma^2\left(\mathrm{Tr}\left(\mathbf{I}\right) - \mathrm{Tr}\left(\mathbf{H}\right)\right) && (15)\\
&= \sigma^2\left(n - p - 1\right) && (16)
\end{aligned}$$

2. Suppose that the conditional distribution of X given that $P = p$ has a binomial distribution wit parameters 5 and p, $X \mid P = p \sim \mathrm{Bin}\left(5, p\right)$ and the marginal distribution of P is a uniform distribution on $(0, 1)$, $P \sim \mathrm{Unif}\left(0, 1\right)$. We would like to calculate the correlation coefficient between X and P.

   (a) Compute variance of X.

   (b) Compute covariance of X and P.

   (c) Compute $\mathrm{Cor}\left(X, P\right)$.

Solution:

3. Let $X_1, X_2, \ldots, X_n \quad (n > 2)$ be a random sample from the following density

$$f\left(x; \theta\right) = \begin{cases} \theta x^{\theta - 1}, & 0 < x < 1,\ 0 < \theta < \infty \\ 0, & \text{elsewhere} \end{cases}$$

   (a) Find the maximum likelihood estimator (MLE) $\widehat{\theta}$ of θ.

   (b) Compare variacne of $\widehat{\theta}$ with the Cramér-Rao lower bound.

Solution:

4. Let $X_1, \ldots, X_n$ be a random sample from the following probability density function(pdf),

$$f\left(x; \theta\right) = \theta / x^2, \quad 0 < \theta \le x < \infty.$$

1. Find a sufficient statistic for θ.
2. Find the maximum likelihood estimator (MLE) of θ.
3. Find the method of moments estimator (MME) of θ.

Solution:

5. Let $X_1, \ldots, X_n$ be a random sample from the following probability density function (pdf),

$$f(x;\theta) = \theta e^{-\theta x}, \quad 0 < x < \infty,$$

where $\theta = \theta_0$ or $\theta = \theta_1$. We assume that known fixed numbers $\theta_1 > \theta_0$.

(a) Explain the *Neyman-Pearson lemma* briefly.

(b) Explain the most powerful (MP) test briefly.

(c) Obtain the MP test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ by the *Neyman-Pearson lemma*.

Solution: