p1

<서식6>

2022년도 후반기 외국어 및 종합시험 답안지

학과별	과정별	시험 과 목 명	수험번호	성 명	감독위원확인
통계	박4	고급회귀분석론	5250	황성윤	

1. (1-1) $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

잔차에 대한 기본가정 : 정규성, 독립성, 등분산성, 선형성.

(1-2) 정규성 : 잔차에 대한 추정값에 대해 Shapiro-Wilk test 등을 통해 검정했을 때 H_0 : normality 를 기각한 경우

등분산성 : 잔차에 대한 산점도의 흩어진 정도가 고르지 않은 경우

독립성, 선형성 : 잔차에 대한 산점도가 특정한 pattern의 형태를 보이는 경우

⇒ 해결방안 : 이상치 제거, 변수변환 등의 방법을 적용해본다.

2. $y_i = \beta x_i + \varepsilon_i$, $i=1, \cdots, n$

(2-1) $Q = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2 \Rightarrow \frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta} x_i) = 0$ $\therefore \hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

(2-2) $\hat{y}_i = \hat{\beta} \cdot x_i = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x_i$ ⇒ 추정된 직선은 (0,0)을 반드시 지나게 된다.

(2-3) $\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$ (항상 만족한다.)

(2-4) $E(\hat{\beta}) = \frac{1}{\sum_{i=1}^{n} x_i^2} E\left(\sum_{i=1}^{n} x_i y_i\right) = \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} x_i \underbrace{E(y_i)}_{\beta x_i} = \frac{1}{\cancel{\sum_{i=1}^{n} x_i^2}} \cdot \beta \cancel{\sum_{i=1}^{n} x_i^2} = \beta$ ($\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ 가정)

$Var(\hat{\beta}) = \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2} Var\left(\sum_{i=1}^{n} x_i y_i\right) = \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2} \sum_{i=1}^{n} x_i^2 \cdot \underbrace{Var(y_i)}_{\sigma^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

<서식7>

p2

<서식6>

2022년도 후반기 외국어 및 종합시험 답안지

학과별	과정별	시험 과 목 명	수험번호	성 명	감독위원확인
통계	박사	고급회귀분석론	5250	황성윤	

3. $\underset{\sim}{y} = X \cdot \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

(3-1) $SSE = \underset{\sim}{e}^t \underset{\sim}{e}$ 의 자유도는 $n-p = n-k-1$ 이다. (k : 설명변수의 개수)

그 이유는 $\underset{\sim}{e}^t \cdot \underset{\sim}{1}_n = \underset{\sim}{0}$, $\underset{\sim}{e}^t \cdot x_j = 0 \ (j=1, \cdots, k)$ 이 성립하기 때문에

총 $k+1$ 개의 제약식이 존재하기 때문이다.

(3-2) $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}.$

$e_i = y_i - \hat{y}_i$

$\underset{\sim}{y}$가 X의 열벡터 중 하나와 일치한다는 것은 결국

반응변수가 k개의 설명변수 중 하나와 일치한다는 것을 뜻한다.

이러한 경우에는 $\underset{\sim}{y}$ 가 X의 열벡터공간 S에 정사영시킨 벡터와

정확히 일치하게 되므로 $\underset{\sim}{e} = \underset{\sim}{0}$ 이 된다.

<서식7>