통계자료분석 9주차 과제

20191245 노유정

<정답>

1번  
실습

2번

1) 타당하다

2) 등분산성이 유지되지 않을 수 있고, 다른 변수들에 대한 고려가 필요하다

3) 약 888만원

4) 약 51만원

5) 잔차가 양의 값을 나타낸다면 예측값이 관찰값보다 크고, 반대로 음의 값이라면 예측값이 관찰값보다 작다는 것을 나타낸다.

1. 21장에 나오는 회귀분석 결과를 R을 이용하여 실행하여 프로그램 코드와 결과를 제출

<문서 마지막에 첨부>

2. 교재 411페이지에 나와 있는 논의 사항 3번:

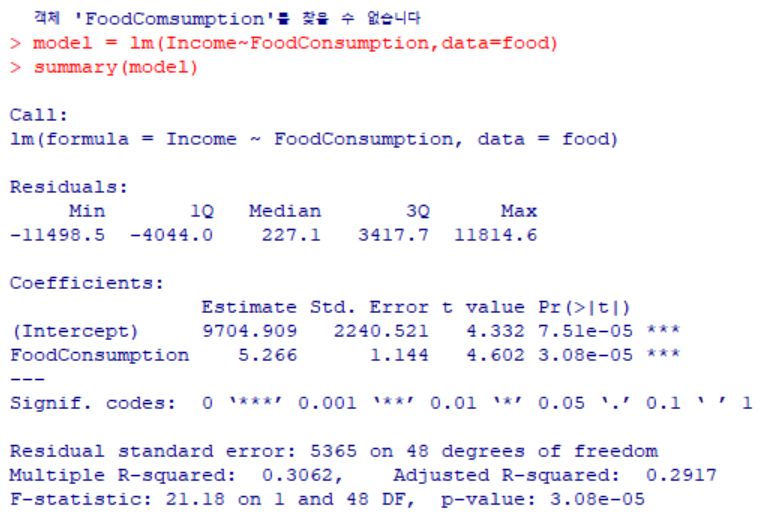
한 가구에서 1년 동안 소비하는 식비지출이 연간소득과 어떠한 관계가 있는지를 파악하기 위하여, 회귀분석을 실시한 결과 다음과 같은 추정된 회귀방정식을 얻을 수 있었으며, 연간소득에 대한 회귀계수는 유의수준 5%에서 유의한 수치로 나타났다.

식비지출 (hat) = 714.05 + 0.0581 (연간소득)

(1) 이러한 회귀방정식이 타당한 지에 대하여 논의하시오. 또한, 연간소득에 대한 회귀계수의 값이 매우 적은 수임에도 불구하고 유의수준 5%에서 유의한 수치가 될 수 있는지에 대하여 논의하시오.

타당하다. 연간소득이 늘어날수록 식비지출이 늘어난다는 계산이 합리적이다

회귀계수의 값은 단위의 조절로 높일 수 있으며, 데이터 분석 결과 p값이 5%보다 작아 유의미한 수치라고 판단할 수 있다.



(2) 문제에서 나타난 내용을 가지고, 회귀모형을 사용하여도 괜찮은지를 판단할 수 있는가? 어떠한 분석이 추가적으로 실시되어야 하는지에 대하여 논의하시오.

연 소득이 다른 가구들의 경우 식대 5만원이 갖는 의미가 서로 다르기 때문에 등분산성이 유지되지 않을 수 있다. 또한 연간소득만이 식비지출에 대한 변수가 아니기 때문에 가구 구성원, 거주지 등 식비에 연관된 변수들에 대한 추가적인 고려가 필요하다.

(3) 3천만원의 연간소득을 갖는 가구가 1년 동안에 소비하는 식비지출은 어느 정도인가?

답: 약 888만원

예측값은 714.05 + 0.0581\*3000 = 888.35(만원)

(4) 4천만원의 연간소득을 갖는 가구가 1백3십만원을 식비지출로 지출하였다면, 잔차는 얼마인가?

답: 약 51만원

예측값은 714.05 + 0.0581\*4000 = 946.45 , 월단위로 78.87, 약 79(만원)임

잔차 = 관찰값 – 예측값 이므로 130 – 79 = 51

(5) 잔차가 양의 값을 나타내는 경우와 음의 값을 나타내는 경우의 차이에 대하여 논의하시오

답: 잔차가 양의 값을 나타낸다면 예측값이 관찰값보다 크고, 반대로 음의 값이라면 예측값이 관찰값보다 작다는 것을 나타낸다.

<1번 실습>

