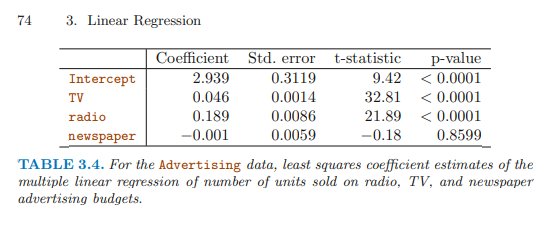
19101185 강동현 데이터분석 Assignment 05

1. 아래의 테이블 3.4를 보고 linear model을 통해 null hypotheses를 설명하고p – value를 통해 어떤 결론을 도출해야하는지 물음에 답하는 문제이다.



H0는 Coefficient가 0인지 판단하는 과정으로 두어 테스팅을 진행한다. 이 경우 신문을 제외한 다른 매체들은 p-value 값이 특별한 값으로 나왔으므로 가설 H0를 수용하지 않는다. 반면 신문의 경우는 p-value 값이 특정 범위 내로 나왔으므로 신문은 H0를 수용 가능하다. 즉 Coefficient에 대해서 0이거나 그 주변의 값이라는 것을 알 수 있고 이는 신문이 전체 매출에 대해서 미치는 영향이 없음을 의미한다.

3. 5개의 변수가 존재한다. 각각 X1 = GPA , X2 = IQ , X3 = 성별 ( 1은 여자 , 0은 남자 ) , X4 = GPA와 IQ 사이의 Interaction , X5 = GPA와 성별 사이의 Interaction 이다. Respone는 졸업 후 초봉이다. Model을 fit 하기 위해서 least squares 를 사용한다. 베타0햇 = 50 , 베타1햇 = 20 , 베타2햇 = 0.07 , 베타3햇 = 35 , 베타4햇 = 0.01 , 베타5햇 = -10 이라고 가정하자.

(a). 어떤 답이 맞는지 답하고 왜 그런지 설명하자. 일단 식을 정리 해보자면

Y = 50 + 20GPA + 0.07IQ + 35성별 + 0.01(GPA \* IQ) - 10 (GPA \* 성별) 이다. 3번의 경우를 살펴보자. IQ과 GPA를 고정하고 남성과 여성의 초봉을 비교하는 것이다. 남성은 0 , 여성은 1을 그대로 대입해 식을 정리해보자.

남자 : Y = 50 + 20GPA + 0.07IQ + 0.01(GPA \* IQ)

여자 : Y = 50 + 20GPA + 0.07IQ + 35 + 0.01(GPA \* IQ) - 10 (GPA)

같은 부분을 제외한 나머지 부분만을 비교해보자면 여자는 남자의 식에 35 -10GPA만 더해진 것을 볼 수 있다. IQ와 GPA는 고정된 값이라고 하였으므로 남자와 여자의 식에 대입되는 IQ과 GPA값 역시 같다고 볼 수 있을 것이다. 따라서 문제에서 GPA가 충분히 높아지는 경우를 제시 했으므로 우리는 35 – 10GPA에서 GPA가 높아질수록 여자의 초봉이 낮아진다는 사실을 알 수 있다. 따라서 정답은 3번이다.

(b). IQ가 110 이고 , GPA가 4.0인 여성의 급여를 예측하는 문제이다. 여자의 식은

Y = 50 + 20GPA + 0.07IQ + 35 + 0.01(GPA \* IQ) - 10 (GPA) 이므로 각각의 값을 대입하면 50 + 80 + 7.7 + 35 + 4.4 – 40 이므로 137.1 로 계산 결과가 나온다. 따라서 여성은 137.1 천 달러의 초봉을 가진다.

(c). 참 거짓을 선택하는 문제이다. 만약 GPA와 IQ의 Interaction이 매우 작다면 , Interaction의 전체에 미치는 영향이 미비하거나 없는지 묻는 문제이다. 일단 우리는 Coefficient가 영향이 있는지 없는지를 판단하기 위해서는 p-value를 확인 해야 한다. 따라서 우리는 p-value를 확인해 통계적으로 H0를 수용가능한지 가능하지 않은지 판단하기 전까지는 Coefficient가 영향을 미치는 지 미치지 않는지 판단할 수 없다. 따라서 우리는 영향을 판단하지 못하므로 정답은 거짓이다.