19101185 강동현 데이터 분석 Assignment 04

1. x와 y 데이터를 보고 물음에 답하는 문제이다.

(a). scatter diagram을 그리는 문제이다.

(b). 다음 데이터들을 가지고 linear curve 만드는 것이다. 따라서 우리는 A , B를 구해야 한다. 먼저 B를 구해야 하므로 n = 8 으로 두고 B값을 구해보자. 계산하면 B값은 약 1.2063 이다. B값을 통해서 A값 역시 구할 수 있다. 식에 따라서 계산해보면 A값은 약 2.4642625 이다. 따라서 식은 y = 2.4642625 + 1.2063x 로 나오게 된다.

2. 가격과 주문량의 관계를 나타낸 데이터들 이다. 이 데이터들을 통해 가격이 25일 때의 주문량을 추정 하는 문제이다. 따라서 A와 B를 통해 추정 할 수 있다. B의 경우 n = 6 인 경우 이므로 계산하면 약 -2.3746 이다. B를 통해서 A를 계산하면 약 206.6940836 이다. 따라서 식은 y = 206.6940836 -2.3746x 이고 x값이 25일때의 y값은 147.3290836 이다. 물건의 주문량은 자연수 이므로 가장 가까운 자연수인 147로 추정 될 것이다. 따라서 가격이 25일 때 데이터들로부터 추정되는 주문량은 147 이다.

3. 자동차의 연료 소비량이 자동차의 주행 속도에 영향을 주지 않는다는 의견을 반박하는 문제이다. 문제에는 주행 속도와 연료 소비량에 대한 데이터들이 주어져 있다.

(a). 두 데이터 간의 선형 관계를 찾는 문제이다. 이 문제 역시 앞선 문제들과 같이 A와 B를 구해야 한다. n = 7 인 경우 이므로 B를 계산하면 B 값은 -0.17 이다. B값을 통해서 A값을 구하면 약 32.54285 이다. 따라서 선형 관계는 y = 32.54285 – 0.17x 로 나오게 된다.

(b). 우리는 데이터들 간에 관계가 성립하는지 보기위해 H0 와 H1을 보면 된다. 일단 SSR을 구해야 한다. SSR의 값은 약 1.5271428575 로 나오게 된다. 또한 SXX의 값은 700으로 구했었다. 따라서 베타에 대해서는 8.13847 라는 v-value 가 나오게 된다. 문제에서는 8.138을 v-value로 두었고 8.13847은 8.138보다 크므로 H0는 틀렸다고 볼 수 있다. 즉 H1을 수용해야 하고 이는 베타에 관해서는 관계성이 입증 된 것이다. 또한 SSR 과 SXX를 이용해 알파에 관해서도 H0 과 H1에 대하여 테스트를 진행하면 된다. A의 경우 v-value는 34.15725816364286 이고 이는 25.612 보다 큰 v-value 이다. 따라서 H1을 수용해야 한다. 베타와 알파에 대해서 모두 관계가 성립한다는 것을 보았으므로 우리는 연료 소비량이 자동차의 주행 속도와 관련이 있다고 결론 내릴 수 있다.

4. y = 베타0 + 베타1 \* x 로 표현 될 수 있다. 이때 다음 식을 보고 점 (x바 , y바 )를 무조건 지나가는지 보이는 문제이다. 식(3.4) 에서 베타1 과 베타0 에 대해 표현 했다. 베타0 의 경우 y바 -베타1 \* x바 이므로 대입하여 식을 정리 할 수 있다. 식을 정리하면 y – y바 = 베타1 ( x – x바) 이다. 이 식에서 x에 x바를 , y에는 y바를 대입하면 베타1에 관계없이 등식이 성립함을 알 수 있다. 따라서 식(3.4)를 통해 simple linear regression의 경우 , 반드시 점 (x바 , y바 )를 지나는 것을 알 수 있다.