

Cars93의 데이터에 대한 요약 정보를 나타내준다. summary의 p-value 값이 지수표기법으로 6.746e-16이므로 0.05보다 훨씬 작기 때문에 유의미하다고 할 수 있다.

=> 따라서 자동차의 가격은 엔진의 크기, RPM, 무게로 추정 가능하다는 뜻이 된다.

[2]

<-1. 회귀분석이란?>

회귀의 뜻을 먼저 살펴보자면 회귀는 한 바퀴 돌아 제자리로 돌아가는 것을 말한다.

즉 데이터 분석에 있어서 데이터들이 제자리(평균값)으로 돌아가려는 성질이 있다는 것을 뜻한다고 할 수 있다.

예를 들어서 < 부모의 키가 크면 자녀의 키도 큰가? >라는 가설을 볼 때 키가 큰 사람들끼리 결혼해서 자녀를 낳는 것을 계속 반복하게 되면 가설이 맞는다면 키가 중첩되어 수십 년이 지나게 되면 인간의 키가 2~3M가 될 텐데 그렇지 않다는 점을 봤을 때 모든 것이 평균으로 돌아간다고 할 수 있다는 것이다.

이처럼 회귀분석은 1개 또는 여러 개의 설명 변인과 하나의 연속적인 반응 변인이 인과적 관계를 갖는지

경향성과 의존성을 알아보고자 할 때 사용하는 분석 방법이다.

[1]의 식을 예로 들자면 독립 변수(엔진 크기, 분당 회전수, 무게)를 통해 종속 변수(가격)를 설명한다고 볼 수 있다.

회귀분석을 사용하는 이유는 위와 같이 독립 변수와 종속 변수 간의 관계를 설명하거나 과거, 현재의 데이터를 빅 데이터, 인공지능, 머신러닝에 접목시켜 미래를 예측하는데 사용하는 것처럼

독립 변수에 의한 종속변수의 예측을 하기 위해 사용된다.

회귀 분석의 종류로는 선형 회귀, 로지스틱 회귀, 푸아송 회귀 등이 있다.

<-2. 회귀분석의 적용 예시>

설명변수(경력, 학력)에 따른 반응변수(연봉)의 관계를 분석하는 것

설명변수(기온)에 따른 반응변수(음료 판매량)의 관계를 분석하는 것

설명변수(키)에 따른 반응변수(몸무게)의 관계를 분석하는 것

등등

<-3. 회귀분석에 대해서 무엇을 공부한지 등의 소감을 적어서 제출한다.>

나는 회귀분석을 검색하고 조사하면서 회귀분석이라는 것이 통계분석에서 가장 자주 사용되는 분석이라는 것을 알 수 있었다.

따라서 이 회귀분석이라는 것을 정확하게 이해하고 공부하게 된다면 나중에 통계를 R 언어를 통해 분석할 때 굉장히 유용할 것이라는 생각이 들었다.

또 회귀분석이라는 것이 선형 회귀와 로지스틱 회귀 등 모두 같은 표현인 줄 알았는데 독립 변수의 개수에 따라 단순 선형 회귀나 다중 선형 회귀나,

다른 데이터 종류의 의존 변수에 따라 로지스틱 회귀(이진 범주형 결과를 모델화), 푸아송 회귀(상수 데이터로 모델화)로 나뉜다는 것을

알 수 있어서 회귀분석을 구체적으로는 모르겠지만 회귀분석에 대한 표현이나 종류가 큰 뿌리로 잡힌다는 느낌이 들었다.

나는 TV에서 항상 개념의 큰 뿌리를 잡고 가지치기를 해나가면 공부가 쉬울 것이라는 뜻이 정확하게 무슨 뜻인지 크게 와닿지 못했다.

하지만 나는 팀 프로젝트로 그래프를 R 명령어들을 통해 꾸미고, taply를 사용하여 t 검정을

한 것과 회귀분석에 대해 스스로 검색하고 조사하며 학습하면서

"아 교수님께서 항상 말씀하시는 큰 뿌리를 핵심으로 가지처럼 뺏어나가는 것이 이거구나"하고 감명 깊게 느낄 수 있었다.

교수님께서 이해가 쉽게 큰 뿌리를 우리에게 가르쳐주시면 우리는 그 큰 뿌리를 복습하고 이해하여 잔가지를 만들어나가는 것이 R 언어 공부에 있어서

효과적이라는 것을 느낄 수 있었던 것이다. 또한 이 큰 뿌리-잔 가지 공부법은 다른 공부를 할 때도 적용시키면 효과적이겠다는 생각이 들었다.

대학에 와서 주체적으로 생활하고 학습하는 것이 무엇인지 이번 통계학개론 수업을 수강하면서 넓고 자세하게 배울 수 있다는 점에서 교수님께 굉장히 감사하다.

앞으로 교수님의 주옥같은 말들을 모두 내 것으로 체화하여 대학을 졸업할 때쯤이 되면 지금보다 더 큰 꿈을 그릴 수 있는 사람이 될 수 있게 노력해야겠다고 생각이 들었다.