

seaborn 패키지의 iris dataset을 가져왔다.

iris 종은 setosa, versicolor, virginica가 존재한다.

각각의 petal\_length를 box plot으로 시각화하였다.

box plot 상에서 petal\_length의 대략적인 길이는 setosa, versicolor, virginica 순으로 컸다.

각각의 데이터가 정규성을 가지는지 Shapiro-Wilk 검정을 통해 확인하였다.

그 결과 3개 종의 petal\_length의 Shapiro-Wilk 검정을 통해 구한 p-value는 모두 0.5 이상, 즉 귀무가설인 "각각의 데이터가 정규분포를 따름"을 확인할 수 있었다.

3개의 데이터가 등분산성을 가지는지 Levene 검정을 통해 확인하였다.

Levene 검정을 통해 얻은 p-value는  $3.12 \cdot 10^{-8}$  정도로, 등분산이라고 보기 어려웠다.

이후 3개 데이터의 petal\_length의 모평균이 같다는 귀무가설을 세우고, 귀무가설을 확인하기 위해 One-way ANOVA 분석을 시행하였다.

시행 결과 p-value는  $2.85 \cdot 10^{-91}$  정도로 귀무가설이 성립하지 않다는 것을 확인할 수 있었다.

3개 데이터의 모분산 차이를 자세히 분석하기 위해 Tukey HSD 사후검정을 실시하였다.

그 결과, box plot을 통해 확인할 수 있었던 setosa, versicolor, virginica 3개 종의 petal\_length 모평균 대소 관계를 다시 확인할 수 있었다.

이후 sklearn 라이브러리를 이용해 회귀 분석을 진행하였다.

$\text{mse}$  값은 0.108,  $R^2$  값은 0.971 정도로 나왔으며, 회귀계수도 구할 수 있었다.  $\text{mse}$ 와  $R^2$  값을 통해 해당 선형 회귀가 주어진 데이터를 꽤나 잘 나타내는 모델이라고 해석하였다.