

1. Boxplot

- setosa, versicolor, virginica 순으로 평균 길이가 긴 것을 확인할 수 있다. 이 때, setosa가 versicolor와 virginica에 비해서 분산이 작아서 비교적 균일한 값을 확인할 수 있다. 또한, setosa와 versicolor에서는 이상치를 관측할 수 있다. 전반적으로 species에 따라서 petal length의 분포 차이가 뚜렷한 것을 확인할 수 있다.

2. 정규성 검정

- H0: 해당 species의 petal_length는 정규분포를 따른다.
- H1: 해당 species의 petal_length는 정규분포를 따르지 않는다.
- setosa, versicolor, virginica 세 개의 species 모두 p-value값이 0.05보다 크기 때문에 귀무가설을 기각하지 않는다. 즉, 세 species의 petal_length 모두 정규 분포를 따름을 알 수 있다.

3. 등분산성 검정

- H0: 세 species의 petal_length는 등분산성을 만족한다.
- H1: 세 species의 petal_length는 등분산성을 만족하지 않는다.
- P-value의 값이 0.05보다 작기 때문에 H0을 기각해야 한다. 즉, 세 species의 petal_length는 등분산성을 만족하지 않는다.

4. One-way ANOVA 검정

- H0: species 간 petal_length의 평균은 모두 동일하다.
- H1: 적어도 하나의 species의 petal_length의 평균은 동일하지 않다.
- One-way ANOVA를 진행하였을 때, p-value의 값은 0.05보다 작은 값으로 나타난다. 즉, H0을 기각해야 하며 적어도 하나의 species의 petal_length의 평균은 동일하지 않다고 할 수 있다.

5. 사후검정 (Tukey HSD)

- One-way ANOVA 이후에 Tukey HSD 사후검정을 수행한 결과, 세 가지 경우의 p-value의 값이 다 0.05보다 작은 것을 알 수 있다. 즉, setosa-versicolor, setosa-virginica, versicolor-virginica의 세 가지의 species의 쌍에서 모두 평균의 차이가 통계적으로 유의하다는 점을 알 수 있다. 이는 결론적으로 species 간 petal_length의 평균이 서로 통계적으로 유의한 차이를 보이는 것을 확인할 수 있다.

6. 결과 요약

- Boxplot를 통해 setosa, versicolor, virginica 순으로 petal_length의 평균 길이가 증가하는 경향을 확인했다. 정규성 검정과 등분산성 검정을 통해 세 species의 petal_length가 모두 정규분포를 따르지만 동일한 분산 값을 가지지는 않는다는 점을 알 수 있었다. One-way ANOVA 검정을 수행한 결과

species 간의 petal_length 평균 값에 통계적으로 유의한 차이가 존재한다는 것을 확인할 수 있었고, 이에 Tukey HSD 사후검정을 시행한 결과, 모든 쌍에서 평균의 유의미한 차이가 존재한다는 것을 알 수 있었다.

7. 회귀분석

- 먼저, 다른 조건이 동일할 때, petal_length의 값은 intercept인 -0.2622지만 이는 모형을 보정해주기 위한 값으로 해석하고자 한다. Sepal_length의 길이가 1만큼 증가할 때 petal_length의 값은 0.7228만큼 증가한다. Sepal_width의 값이 1만큼 증가할 때 petal_length의 값은 0.6358만큼 감소한다. Petal_width의 값이 1만큼 증가할 때, petal_length의 값은 1.4675만큼 증가한다. MSE의 값이 0.1300으로 이는 예측 오차가 크지 않음을 의미한다. R^2 의 값은 0.9603으로 이 모델의 설명력이 매우 높음을 알 수 있다. 전반적으로 해당 회귀모형의 성능은 매우 우수하다고 판단할 수 있다.