

Report

3(1) Basic_Statistics - Report

2021147586 안재후

3. 시각화

- Setosa 종은 꽃잎의 길이가 1cm ~ 2cm 사이에 분포하고 있고, Versicolor 종과 Virginica 종에 비해 현저히 짧고 길이 분포도 좁다.
- Virginica 종은 꽃잎의 길이가 다른 두 종에 비해 가장 크고, 분포도 가장 넓다.
- Versicolor와 Virginica는 일부 겹치는 구간이 존재하지만, 전체적으로 Virginica의 값이 더 크고 사분위수(box)의 범위가 겹치지 않는다. 따라서 꽃잎 길이만으로 종을 어느 정도 구별할 수 있다.

4. 정규성 검정

가설 설정

H_0 : 해당 종의 Petal Length 데이터는 정규분포를 따른다.

H_1 : 해당 종의 Petal Length 데이터는 정규분포를 따르지 않는다.

5. 등분산성 검정

Levene 검정을 통한 3개 그룹 간 등분산성 검정

H_0 : 모든 그룹의 분산이 동일하다.

H_1 : 적어도 한 그룹의 분산이 다르다.

6. ANOVA 가설 수립

H_0 : 세 species 간 sepal_length의 평균이 동일하다.

H_1 : 적어도 한 species의 sepal_length의 평균이 다르다.

7. One-way ANOVA

귀무가설을 기각한다. 즉, 세 품종 간 Petal Length 평균에는 통계적으로 유의미한 차이가 존재한다.

8. 사후검정

모든 종 간의 유의미한 차이($p\text{-adj} < 0.05$)가 있는지 확인한 결과, 모든 종에서 귀무가설이 기각되었다.

9. 결과 요약

Boxplot, ANOVA, 사후검정 결과를 종합한 결과, 세 품종(Setosa, Versicolor, Virginica) 간의 꽃잎 길이의 평균은 통계적으로 유의미한 차이가 있음을 확인하였다. 이는 시각화에서 시각적으로 분석한 바와 동일하다. 특히 Setosa의 Petal Length가 가장 짧으며, Virginica가 가장 긴 것으로 보인다. Tukey HSD를 통한 사후검정 결과 One-Way ANOVA의 결과가 검증되어 모든 품종 쌍 사이에서 유의미한 평균 차이가 존재함을 알 수 있다.

10. 회귀 분석

- R^2 값이 0.99로 1에 매우 가까워, 모델이 petal_length의 변동을 매우 잘 설명하고 있다.
- 회귀계수를 보면 petal_width가 1.45로 가장 큰 영향을 미치며, sepal_length 또한 양의 상관관계를 보인다.
- MSE가 0.04로 매우 작다 (예측 오차가 적다).
- 전체적으로 꽃잎의 길이와 꽃잎의 너비가 양의 상관관계를 보여, 앞서 분석한 꽃잎 길이의 분포와 더불어 분석할 수 있다고 판단된다. 꽃받침의 길이 또한 꽃잎의 길이와 양의 상관관계를 보인다.