

[3회차] 기초통계 과제

🕒 Created	@2026년 1월 19일 오후 2:36
🏷 Tags	

시각화

boxplot을 통해 species별 petal length의 분포를 비교한 결과

1. setosa는 petal length의 값이 가장 작고 분산이 작아 비교적 균일한 분포를 보였다.
2. versicolor는 setosa보다 큰 값을 가지고 중간 수준의 분포를 보였다.
3. virginica는 petal length의 중앙값과 범위가 가장 커 가장 큰 변동성을 보였다.

이를 통해 petal length는 species를 구분하는 데 유용한 변수임을 확인할 수 있다.

정규성 검정 (Shapiro-Wilk)

1. 귀무가설 (H_0): 해당 species의 petal length 데이터는 정규분포를 따른다.
2. 대립가설 (H_1): 해당 species의 petal length 데이터는 정규분포를 따르지 않는다.

species별 petal length에 대해 Shapiro-Wilk 정규성 검정을 실시한 결과, setosa($p=0.0548$), versicolor($p=0.1585$), virginica($p=0.1098$) 모두 유의수준 0.05보다 큰 p-value를 보였다.

따라서 각 species의 petal length는 정규성을 만족한다고 가정할 수 있으며 이후 분석에서는 정규성을 가정하였다.

등분산성 검정 (Levene)

1. 귀무가설 (H_0) : 세 species 그룹의 petal length 분산은 서로 같다.
2. 대립가설 (H_1) : 적어도 한 species 그룹의 petal length 분산은 다르다.

species별 petal length의 등분산성을 Levene 검정을 통해 검정한 결과, p-value는 0.0000으로 0.05보다 작은 값으로 나타나 귀무가설을 기각하였다. 따라서 세 그룹 간 분산

이 동일하다고 보기 어렵다.

ANOVA 가설 수립

1. 귀무가설 (H_0) : 세 species 간 petal length의 평균은 모두 같다.
2. 대립가설 (H_1) : 적어도 하나의 species는 petal length의 평균이 다르다.

One-way ANOVA

유의수준 0.05에서 One-way ANOVA 결과 p-value가 0.05보다 작으므로 귀무가설을 기각하며, species 간 petal length 평균에는 통계적으로 유의한 차이가 존재한다.

사후검정 (Tukey HSD)

Tukey HSD 사후검정 결과 setosa-versicolor, setosa-virginica, versicolor-virginica 모든 종 쌍에서 보정된 p-value가 유의수준 0.05보다 작게 나타나 귀무가설을 기각하였다. 즉, 모든 species 간 petal length 평균에는 통계적으로 유의한 차이가 존재한다.

결과 요약

Boxplot, One-way ANOVA, Tukey HSD 사후검정 결과를 종합하면 petal length는 setosa가 가장 짧고, versicolor가 중간, virginica가 가장 긴 것으로 나타났으며 세 그룹 간 차이는 모두 통계적으로 유의하였다.

회귀 분석

sepal_length, sepal_width, petal_width를 입력 변수로 한 선형회귀 분석 결과, mse가 0.143으로 작게 나타나 예측 오차가 크지 않다는 것을 알 수 있다.

R^2 값이 0.957로 높게 나타나 모델의 설명력이 매우 우수하다는 것을 확인하였다. 회귀계수 해석 결과 petal_width가 petal_length에 가장 큰 양의 영향을 미쳤으며, sepal_length는 양의 영향, sepal_width는 음의 영향을 주는 것으로 나타났다.