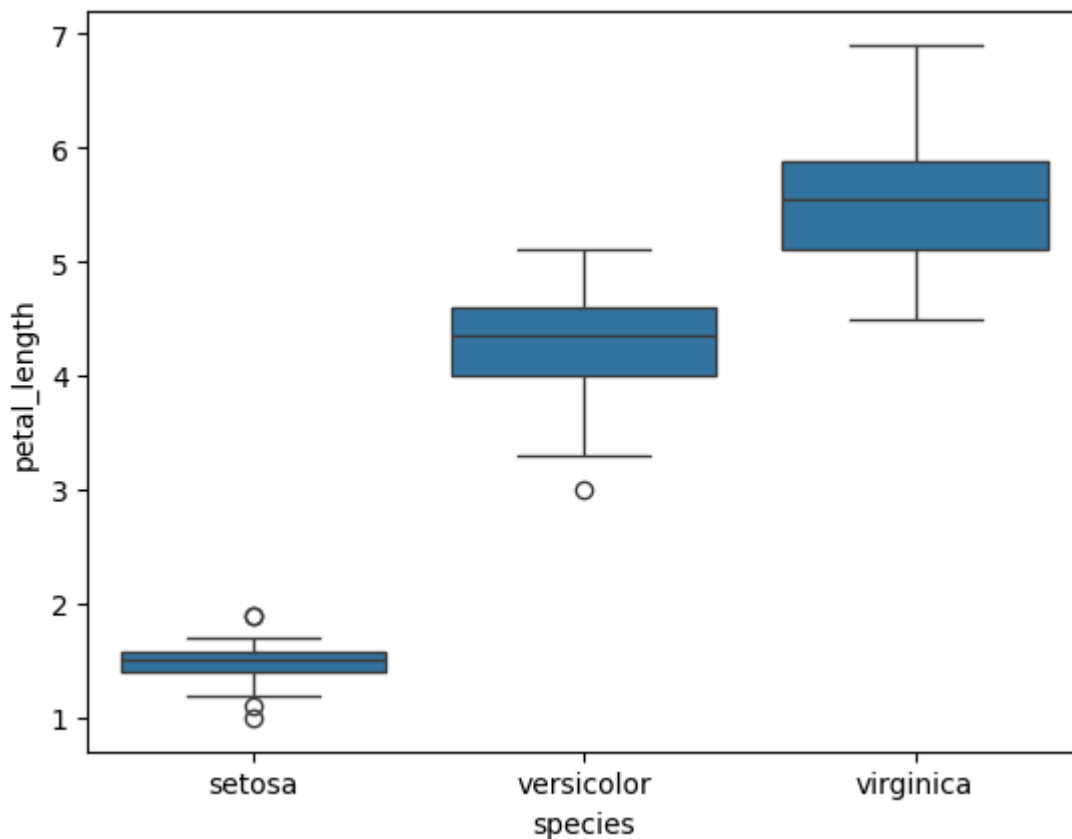


와빅 기초통계 과제

3번



Boxplot 결과, setosa는 petal_length가 전반적으로 매우 짧고 분산도 작아 값이 비교적 일정한 반면, versicolor는 setosa보다 뚜렷하게 큰 값들을 가지며 분산이 더 큰 편이다. virginica는 세 종 중 petal_length의 중앙값과 전체 분포가 가장 높게 나타나, 꽃잎 길이가 가장 긴 종임을 보여준다. 전반적으로 세 종의 분포가 명확히 구분되어 있어 species에 따라 petal_length 평균에 차이가 있을 가능성이 크다.

4번

가설 수립

H_0 : 해당 species의 petal_length는 정규분포를 따른다.

H_1 : 해당 species의 petal_length는 정규분포를 따르지 않는다.

Species별 p-value

setosa: 0.0548114671955363

versicolor: 0.15847783815657573

virginica: 0.10977536903223506

해석

귀무가설 하에서, 모든 species에서 p-value > 0.05이므로 귀무가설을 기각하지 않는다. 따라서 유의수준 0.05에서, 각 species의 petal_length는 정규성을 만족한다고 볼 수 있다.

5번

가설 수립

H_0 : 세 species의 petal_length 분산은 같다

H_1 : 적어도 한 그룹의 분산이 다르다

p-value

3.1287566394085397e-08

해석

귀무가설 하에서, p-value < 0.05이므로 귀무가설을 기각한다. 따라서 유의수준 0.05에서, 적어도 한 그룹의 분산이 다르다고 볼 수 있다.

6번

H_0 : 세 species 간 petal_length의 평균은 모두 같다.

H_1 : 적어도 한 species의 petal_length 평균은 서로 다르다.

7번

- **F-value:** 1180.16
- **p-value:** 2.86×10^{-91}

귀무가설 하에서, p-value < 0.05이므로 귀무가설을 기각한다.

즉, 유의수준 0.05에서, 세 species 간 petal_length의 평균은 모두 같다고 볼 수 없으며, 적어도 한 종의 평균 petal_length가 다른 종과 통계적으로 유의하게 다르다고 볼 수 있다.

8번

Anova 결과가 유의하므로, TukeyHSD를 실행한 결과는 다음과 같다.

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
setosa	versicolor	2.798	0.0	2.5942	3.0018	True
setosa	virginica	4.09	0.0	3.8862	4.2938	True
versicolor	virginica	1.292	0.0	1.0882	1.4958	True

Tukey HSD 사후검정 결과, 모든 종 쌍에서 petal_length 평균 차이가 통계적으로 유의하다.

- **setosa – versicolor**

평균 차이(meandiff) = **2.798**, $p\text{-adj} < 0.05 \rightarrow$ **유의한 차이 있음**

- **setosa – virginica**

평균 차이(meandiff) = **4.09**, $p\text{-adj} < 0.05 \rightarrow$ **유의한 차이 있음**

- **versicolor – virginica**

평균 차이(meandiff) = **1.292**, $p\text{-adj} < 0.05 \rightarrow$ **유의한 차이 있음**

또한 모든 비교에서 신뢰구간이 0을 포함하지 않고, reject = True로 나타나 귀무가설을 기각할 수 있다.

따라서 세 species는 서로 명확히 구분되는 평균 차이를 가진다.

9번

Boxplot 결과, setosa의 petal_length가 가장 짧고 versicolor는 중간, virginica가 가장 길게 나타나 species별 분포 차이가 뚜렷했다. One-way ANOVA 결과 p-value가 0.05 보다 매우 작아 세 species의 petal_length 평균이 같다는 귀무가설은 기각되었으며, 종에 따른 평균 차이가 통계적으로 유의함을 확인하였다. Tukey HSD 사후검정에서도 모든 종 쌍에서 유의한 평균 차이가 나타났고, petal_length는 setosa < versicolor < virginica 순으로 유의하게 증가하는 것으로 결론지을 수 있다.

10번

MSE: 0.1300162603138268,

R^2 : 0.9603293155857664,

회귀 계수: [0.72281463, -0.63581649, 1.46752403]

테스트 데이터 기준, MSE는 0.130으로, 예측값이 실제 petal_length와 비교적 작은 오차를 보인다. 결정계수는 0.96으로, 입력 변수들이 petal_length 변동의 약 96%를 설명하는 매우 높은 설명력을 가진다. 회귀계수를 보면 sepal_length(0.723)와 petal_width(1.468)는 petal_length에 양의 영향을 미치며, petal_width의 영향이 가장 크다. 반면 sepal_width(-0.636)는 다른 변수들이 고정될 때 petal_length에 음의 영향을 미친다.