

day3-가설검증

1번 문제

시나리오: 어떤 게임 개발사에서 새로운 아이템의 뽑기 확률이 10%로 설정되었다고 주장하고 있습니다. 유저 커뮤니티에서는 "실제 확률은 10%보다 낮은 것 같다"는 의혹이 제기되었습니다. 이를 확인하기 위해 한 유저가 아이템을 200번 뽑았고, 그중 12번 성공했습니다.

수행 과제: 이항검정(Binomial Test)을 사용하여, 이 유저의 결과(200번 중 12번 성공)가 개발사의 주장($P=0.1$)을 기각하고 "확률이 10%보다 낮은 것 같다"는 의혹을 뒷받침할 만큼 통계적으로 유의미한지 검증하세요.

*이항검정이란

두 개의 범주(예: 성공/실패, 만족/불만족 등)로 구분된 이항분포 데이터에서 모집단의 비율(비)이 특정 값과 같은지, 다른지를 통계적으로 검정하는 방법이다.

(가설 설정: 귀무가설(H_0)은 모집단 비율이 특정 값과 같다는 것이고, 대립가설(H_1)은 그 값과 다르다는 것입니다. 검정 방향(양측/단측)에 따라 가설이 달라질 수 있습니다)

```
from scipy.stats import binomtest

# 문제 설정
n = 200 # 총 시행 횟수
x = 12 # 관찰된 성공 횟수
p = 0.1 # 귀무가설에서의 성공 확률

# [작성] 가설을 주석으로 작성해보세요.
# H0 (귀무가설): 아이템이 나올 확률은 10%이다. (p = 0.1)
# H1 (대립가설): 아이템이 나올 확률은 10%보다 낮다. (p < 0.1)

# [작성] 이항검정을 수행하고 p-값을 계산하세요.
# '확률이 더 낮은 것 같다'는 주장을 검증하려면 alternative 인수를 어떻게 설정해야 할까요?

# [이항검정 수행]
result = binomtest(12, 200, 0.1, alternative='less')
p_value = result.pvalue

print(f"관찰 결과: {n}번 시도 중 {x}번 성공")
print(f"검정 결과 p-value: {p_value:.4f}")

# [유의수준 기준]
alpha = 0.05

if p_value < alpha:
    print("결론: p-value < 0.05 이므로 귀무가설을 기각합니다.")
    print("→ 아이템 등장 확률이 10%보다 낮다는 주장은 통계적으로 유의미합니다.")
else:
    print("결론: p-value ≥ 0.05 이므로 귀무가설을 기각할 수 없습니다.")
    print("→ 아이템 등장 확률이 10%보다 낮다고 말할 충분한 근거가 없습니다.")
```

최종결과

관찰 결과: 200번 시도 중 12번 성공
검정 결과 p-value: 0.0320
결론: p-value < 0.05 이므로 귀무가설을 기각합니다.
→ 아이템 등장 확률이 10%보다 낮다는 주장은 통계적으로 유의미합니다.

alternative= 'less' 쓴 이유 가설 내용의 기준은 **10%**이다. 그 중 우리는 **왼쪽(더 낮은 확률)**만 관심 있음. 따라서 오른쪽 꼬리는 버리고 **×** 왼쪽 꼬리만 본다 **O**

👉 이게 바로 **단측검정**이고 그 코드 표현이 `alternative='less'`

1-1문제

수행 과제: Part 1에서 계산된 p-값의 의미를 직접 눈으로 확인해 봅시다. "만약 개발사의 주장대로 실제 확률이 정말 10%라면", 200번 뽑기를 했을 때 성공 횟수가 어떻게 분포하는지 시뮬레이션을 통해 알아보고, 우리가 관찰한 '12번 성공'이 얼마나 희귀한 일인지 확인하세요.

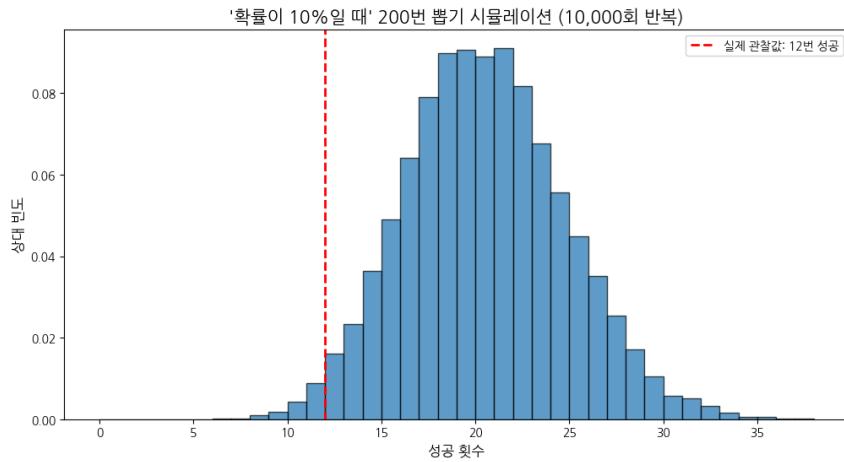
```
# 1. 시뮬레이션 설정
# H0가 사실이라고 가정: 실제 확률 p = 0.1
# 200번 뽑기를 10,000번 반복 수행
num_simulations = 10000

# [작성] numpy의 이항분포 랜덤 함수를 사용하여 시뮬레이션 결과를 생성하세요.
simulated_successes = np.random.binomial(
    n=n,
    p=p,
    size=num_simulations
)

# 2. 시각화
plt.figure(figsize=(12, 6))
# [작성] 시뮬레이션 결과(simulated_successes)를 히스토그램으로 그리세요.
plt.hist(
    simulated_successes,
    bins=range(0, max(simulated_successes) + 2),
    density=True,
    alpha=0.7,
    edgecolor='black'
)

# 실제 관찰값 표시
plt.axvline(
    x=x,
    color='red',
    linestyle='--',
    linewidth=2,
    label=f'실제 관찰값: {x}번 성공'
)

plt.title("확률이 10%일 때 200번 뽑기 시뮬레이션 (10,000회 반복)", fontsize=15)
plt.xlabel("성공 횟수", fontsize=12)
plt.ylabel("상대 빈도", fontsize=12)
plt.legend()
plt.show()
```



① Part 1에서 구한 p-value는 그래프의 어느 부분이며, 의미하는 것

- 그래프에서 빨간 점선(12회 성공)보다 왼쪽 영역 전체가 바로 p-value에 해당하는 부분이다.
- 의미하는 것:

"만약 실제 확률이 정말 10%라면,
200번 뽑기에서 12회 이하로 성공할 확률"

즉, p-value는 이 왼쪽 꼬리 면적이다.

② '12회 성공'은 개발사 주장이 맞다는 가정 하에 흔한가요, 드문가요?

- 드문 일이다.
- 이유:
 - 성공 횟수는 대부분 **20회** 전후에 몰려 있음
 - 12회는 분포의 **왼쪽 끝(꼬리)**에 위치
- 해석:

확률이 10%라면 이렇게 적게 나오는 경우는
자주 발생하지 않는다

③ "p-value가 작으면 귀무가설을 기각한다"를 친구에게 쉽게 설명하면?

"게임사가 말한 확률이 진짜라면 이런 결과가 나올 가능성성이 너무 낮음
그래서 '우연'이라고 보기 어렵고, 개발사 말이 틀렸을 수도 있다고 의심"

p-value가 작다는 건
'이 결과를 우연이라고 믿기엔 너무 억지'라는 뜻이다.

💡 한 줄 요약

- p-value = 그래프에서 관찰값보다 극단적인 영역
- 12회 성공 = 아주 드문 결과
- 그래서 → 귀무가설(확률 10%)을 기각

2번문제

시나리오: 한 식품 공장에서 생산하는 과자 한 봉지의 목표 중량은 150g입니다. 품질관리팀은 생산 공정이 목표 중량을 잘 맞추고 있는지 확인하기 위해, 생산된 과자 30봉지를 무작위로 추출하여 무게를 측정했습니다. (측정 데이터는 아래 코드에 제공)

수행 과제: 일표본 t-검정(One-sample t-Test)을 사용하여, 추출된 과자 샘플의 평균 무게가 목표 중량 150g과 통계적으로 유의미한 차이가 있는지 검증하세요.

```
import numpy as np
from scipy.stats import ttest_1samp

# 주어진 데이터
np.random.seed(42)
sample_weights = np.random.normal(loc=151.5, scale=2, size=30)
pop_mean = 150 # 목표 중량 (모평균)

# [작성] 가설을 주석으로 작성해보세요.
# H0 (귀무가설): 과자 한 봉지의 평균 무게는 150g이다.
# H1 (대립가설): 과자 한 봉지의 평균 무게는 150g과 다르다.

# [작성] 일표본 t-검정을 수행하여 t-통계량과 p-값을 구하세요.
t_statistic, p_value = ttest_1samp(sample_weights, pop_mean)

print(f"샘플 평균 무게: {np.mean(sample_weights):.2f}g")
print(f"t-statistic: {t_statistic:.4f}")
print(f"p-value: {p_value:.4f}")

# [유의수준 0.05 기준 결론]
alpha = 0.05

if p_value < alpha:
    print("결론: p-value < 0.05 이므로 귀무가설을 기각합니다.")
    print("→ 과자 평균 무게는 목표 중량 150g과 통계적으로 유의미한 차이가 있습니다.")
else:
    print("결론: p-value ≥ 0.05 이므로 귀무가설을 기각할 수 없습니다.")
    print("→ 과자 평균 무게는 목표 중량 150g과 유의미한 차이가 있다고 보기 어렵습니다.")
```

결과:

```
샘플 평균 무게: 151.12g
t-statistic: 3.4193
p-value: 0.0019
결론: p-value < 0.05 이므로 귀무가설을 기각합니다.
→ 과자 평균 무게는 목표 중량 150g과 통계적으로 유의미한 차이가 있습니다.
```

2-1문제

수행 과제: 귀무가설이 사실일 때, 즉 "과자 봉지의 평균 무게가 정확히 150g이라면" 어떤 결과들이 나타날 수 있는지 시뮬레이션으로 확인해봅시다. 이를 통해 우리가 실제로 관찰한 샘플 평균(약 151.5g)이 얼마나 특이한 값인지 직접 확인해 보세요.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# 기존 샘플 (이미 앞 문제에서 생성됨)
np.random.seed(42)
sample_weights = np.random.normal(loc=151.5, scale=2, size=30)

# 시뮬레이션 설정
num_simulations = 10000
sample_size = 30
```

```

simulated_means = []

# [시뮬레이션 수행]
# H0: 평균이 150g인 모집단에서 표본을 뽑는 상황을 가정
population_std = np.std(sample_weights, ddof=1) # 샘플 표준편차 사용

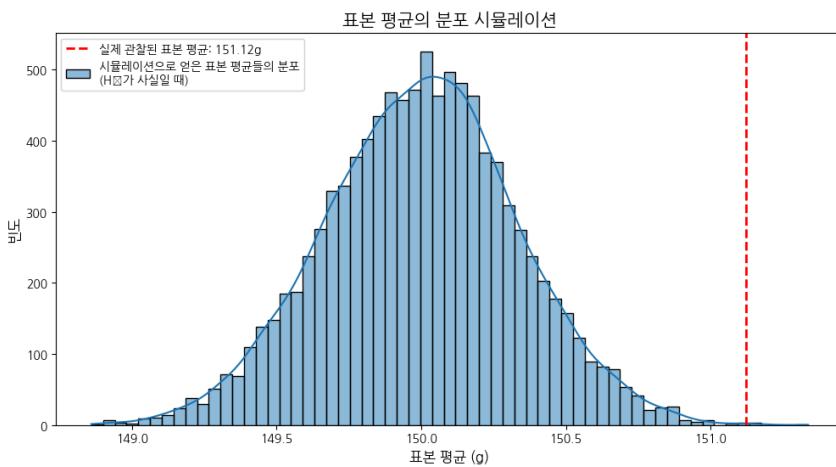
for _ in range(num_simulations):
    simulated_sample = np.random.normal(
        loc=150,           # 귀무가설에서의 평균
        scale=population_std, # 샘플에서 얻은 표준편차
        size=sample_size     # 표본 크기 30
    )
    simulated_means.append(np.mean(simulated_sample))

# 2. 시각화
plt.figure(figsize=(12, 6))
sns.histplot(
    simulated_means,
    kde=True,
    label='시뮬레이션으로 얻은 표본 평균들의 분포\n(H0가 사실일 때)'
)

plt.axvline(
    x=np.mean(sample_weights),
    color='red',
    linestyle='--',
    linewidth=2,
    label=f'실제 관찰된 표본 평균: {np.mean(sample_weights):.2f}g'
)

plt.title("표본 평균의 분포 시뮬레이션", fontsize=15)
plt.xlabel("표본 평균 (g)", fontsize=12)
plt.ylabel("빈도", fontsize=12)
plt.legend()
plt.show()

```



1) Part 2의 히스토그램은 무엇을 나타내나요? 왜 정규분포와 비슷한가요?

답안

Part 2의 히스토그램은 귀무가설이 사실일 때, 즉 과자 봉지의 실제 평균 무게가 150g이라고 가정했을 때

표본 크기 30인 샘플을 반복 추출하여 계산한 표본 평균들의 분포를 나타낸다.

이 분포가 정규분포와 비슷한 모양을 보이는 이유는 **중심극한정리(Central Limit Theorem)** 때문이다.

모집단의 분포 형태와 무관하게, 표본의 크기가 충분히 크면 표본 평균의 분포는 정규분포에 가까워진다.

2) 실제 관찰된 표본 평균은 흔한 값인가요? p-값과 어떤 관련이 있나요?

답안

실제로 관찰한 표본 평균(빨간 점선)은 히스토그램 분포의 중심이 아닌 **꼬리 부분**에 위치하며,

귀무가설이 사실일 때 **흔하게 나타나는 값은 아니다**. Part 1에서 계산된 p-값은 이 분포에서 실제 관찰된 표본 평균보다 같거나 더 극단적인 값이 나올 확률을 의미한다. 따라서 p-값이 작다는 것은 그래프에서 빨간 점선이 분포의 꼬리 영역에 위치함을 의미한다.

3) p-값이 0.001이었다면 빨간 점선은 어디에 있을까요? 공장은 무엇을 해야 할까요?

p-값이 0.001이었다면, 빨간 점선은 히스토그램의 **아주 끝단**, 거의 관측되지 않는 극단적인 꼬리 영역에

위치할 것으로 예상된다. 이는 현재의 생산 결과가 '우연'으로 발생했다고 보기 매우 어렵다는 의미이므로, 공장 입장에서는 생산 공정에 체계적인 문제가 있다고 판단하고 설비 조정, 공정 점검 등 **즉각적인 품질 관리 조치**를 취해야 함을 시사한다.

문제 3

시나리오: 한 교육 기업에서 두 가지 다른 온라인 학습 방식(A, B)을 개발했습니다. 방식 A가 방식 B보다 학생들의 성적 향상에 더 효과적인지 알아보기 위해, 두 그룹의 학생들에게 각각 다른 방식으로 한 달간 학습시킨 후 시험을 보게 했습니다.

수행 과제: 이표본 t-검정(Two-sample t-Test)을 수행하여 두 학습 방식 간 평균 점수 차이가 통계적으로 유의미한지 확인하세요. (단, t-검정의 기본 가정인 정규성과 등분산성을 먼저 검증해야 합니다.)

```
import numpy as np
from scipy.stats import shapiro, levene, ttest_ind

# 데이터 생성
np.random.seed(0)
group_a_scores = np.random.normal(loc=85, scale=8, size=50)
group_b_scores = np.random.normal(loc=80, scale=7, size=50)

# -----
# 1. 정규성 검정 (Shapiro-Wilk test)
# H0: 데이터는 정규분포를 따른다.
# -----
shapiro_a_pvalue = shapiro(group_a_scores).pvalue
shapiro_b_pvalue = shapiro(group_b_scores).pvalue

print(f"A그룹 정규성 검정 p-value: {shapiro_a_pvalue:.4f}")
print(f"B그룹 정규성 검정 p-value: {shapiro_b_pvalue:.4f}")

# -----
# 2. 등분산성 검정 (Levene's test)
# H0: 두 그룹의 분산은 같다.
# -----
levene_pvalue = levene(group_a_scores, group_b_scores).pvalue
print(f"등분산성 검정 p-value: {levene_pvalue:.4f}")

# -----
# 3. 이표본 t-검정
# H0: 두 그룹의 평균은 같다.
# H1: 두 그룹의 평균은 다르다.
# -----
# 등분산성 검정 결과에 따라 equal_var 설정
```

```

equal_var = True if levene_pvalue >= 0.05 else False

t_statistic, p_value = ttest_ind(
    group_a_scores,
    group_b_scores,
    equal_var=equal_var
)

print(f"\n이표본 t-검정 t-statistic: {t_statistic:.4f}")
print(f"이표본 t-검정 p-value: {p_value:.4f}")

```

최종결과

A그룹 정규성 검정 p-value: 0.8766
B그룹 정규성 검정 p-value: 0.8366
등분산성 검정 p-value: 0.0150
이표본 t-검정 t-statistic: 4.0426
이표본 t-검정 p-value: 0.0001
⇒정규성 및 등분산성 검정을 통해 t-검정의 가정을 확인한 후 이표본 t-검정을 수행한 결과,
p-value가 유의수준 0.05보다 작아 두 학습 방식 간 평균 점수 차이는 통계적으로 유의미하다고 판단된다.

2-2

귀무가설, 즉 "두 학습 방식의 효과가 완전히 똑같다면" 어떤 결과가 나타날지 시뮬레이션으로 확인해봅시다. 두 그룹의 점수 차이가 우연히 발생할 수 있는 범위를 눈으로 보고, 우리가 관찰한 실제 점수 차이가 그 범위 안에 있는지 확인해 보세요.

```

import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# 1. 시뮬레이션 설정
# H0가 사실이라고 가정: 두 그룹의 평균은 같다.
# 두 그룹의 모든 데이터를 합쳐서 하나의 거대한 모집단(H0)을 만듭니다.
combined_scores = np.concatenate([group_a_scores, group_b_scores])
num_simulations = 10000
simulated_diffs = []

# [작성] for 반복문을 사용하여 시뮬레이션을 10,000번 수행하세요.
# 각 반복마다, 합쳐진 데이터(combined_scores)에서 비복원추출로 50개(가상 A그룹)와 나머지 50개(가상 B그룹)를 뽑아
# 두 그룹의 평균 차이를 계산하고 simulated_diffs 리스트에 추가하세요.
for _ in range(num_simulations):

    # 합쳐진 데이터에서 비복원추출로 100개를 무작위로 섞음
    shuffled = np.random.permutation(combined_scores)

    # 앞 50개를 가상 A그룹, 뒤 50개를 가상 B그룹으로 나눔
    simulated_a = shuffled[:50]
    simulated_b = shuffled[50:]

    # 두 그룹 평균 차이 계산
    diff = np.mean(simulated_a) - np.mean(simulated_b)
    simulated_diffs.append(diff)

# 2. 시각화
observed_diff = np.mean(group_a_scores) - np.mean(group_b_scores)

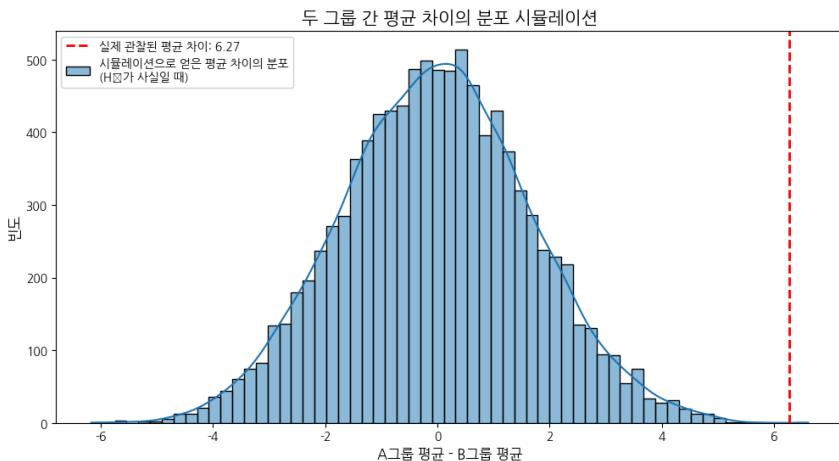
```

```

plt.figure(figsize=(12, 6))
sns.histplot(
    simulated_diffs,
    kde=True,
    label='시뮬레이션으로 얻은 평균 차이의 분포\\n(H0가 사실일 때)'
)
plt.axvline(
    x=observed_diff,
    color='red',
    linestyle='--',
    linewidth=2,
    label=f'실제 관찰된 평균 차이: {observed_diff:.2f}'
)

plt.title("두 그룹 간 평균 차이의 분포 시뮬레이션", fontsize=15)
plt.xlabel("A그룹 평균 - B그룹 평균", fontsize=12)
plt.ylabel("빈도", fontsize=12)
plt.legend()
plt.show()

```



1) 정규성 검정과 등분산성 검정의 p-값은 각각 어떻게 해석해야 하나요? 이 결과는 우리가 이표본 t-검정을 사용하는 데 문제가 없음을 보여주나요?

- 정규성 검정(Shapiro-Wilk)
 - p-value $\geq 0.05 \rightarrow$ “정규분포라고 봐도 무방” (정규성 가정 위반 증거 없음)
 - p-value $< 0.05 \rightarrow$ “정규분포가 아닐 수 있음” (가정 위반 의심)
- 등분산성 검정(Levene)
 - p-value $\geq 0.05 \rightarrow$ 두 집단 분산이 같다고 볼 수 있음 $\rightarrow \text{equal_var=True}$ (일반 Student t-test)
 - p-value $< 0.05 \rightarrow$ 분산이 다를 수 있음 $\rightarrow \text{equal_var=False}$ (Welch t-test 권장)

✓ 결론:

정규성이 크게 문제 없고, 등분산성이 깨졌더라도 Welch t-검정으로 진행하면 이표본 t-검정을 쓰는 데 문제 없다. 즉 “t-검정 자체가 불가”가 아니라, 등분산 여부에 따라 t-검정 종류를 바꾸면 된다는 뜻이야.

2) Part 2 히스토그램 중심이 0에 가까운 이유

- Part 2는 귀무가설(H_0 : 두 방식 효과가 같다, 즉 두 그룹 평균이 같다)이 참이라고 가정한 시뮬레이션이야.
- 효과가 같다면, A/B를 어떻게 섞어 나눠도 평균 차이(A-B)의 기대값은 0이 된다.

👉 그래서 시뮬레이션으로 만든 "평균 차이 분포"는 0을 중심으로 좌우 대칭에 가깝게 나오는 게 정상이다.

3) 실제 관찰된 평균 차이(빨간 점선)는 "두 방식의 효과가 같다"고 가정했을 때 우연히 나타날 수 있는 범위 안에 있나요, 아니면 그 범위를 벗어나나요? 이 시각적 판단과 Part 1의 p-값은 어떤 관계가 있나요?

- 그래프에서 빨간 점선(약 6.27)은 분포의 대부분이 모여 있는 구간이 아니라 거의 끝(극단적인 꼬리)에 있음. 즉, "두 방식 효과가 같다"는 가정 하에서는 우연히 나오기 매우 어려운 값임.

👉 **p-value**가 작다 = "귀무가설이 참일 때, 이런(또는 더 극단적인) 차이가 나올 확률이 작다"

그래프에서는 그게 빨간 점선이 꼬리에 박혀 있는 모습으로 나타난다.

문제 5

시나리오: 한 농업 연구소에서 새로 개발한 비료 3종류(A, B, C)의 생산량 증대 효과를 비교하고자 합니다. 동일한 조건의 밭 30개를 준비하여, 각각 10개씩 비료 A, B, C를 투여한 후 수확량을 측정했습니다.

수행 과정:

- 분산분석(ANOVA)을 사용하여 세 비료 간 수확량 평균에 유의미한 차이가 있는지 검증하세요.
- 만약 ANOVA 검정 결과가 유의미하다면, 어떤 비료가 다른 비료와 차이를 보이는지 확인하기 위해 사후분석(Post-hoc test)인 **Tukey's HSD** 검정을 수행하고 결과를 해석하세요.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# 데이터 생성
np.random.seed(123)
fertilizer_a = np.random.normal(loc=105, scale=10, size=10)
fertilizer_b = np.random.normal(loc=120, scale=12, size=10)
fertilizer_c = np.random.normal(loc=108, scale=9, size=10)

# 데이터프레임으로 변환 (사후분석을 위해)
df = pd.DataFrame({'수확량': np.concatenate([fertilizer_a, fertilizer_b, fertilizer_c]),
                   '비료종류': ['A']*10 + ['B']*10 + ['C']*10})

# -----
# 1. 시각화 (탐색적 분석)
# -----
sns.boxplot(x='비료종류', y='수확량', data=df)
plt.title('비료 종류에 따른 수확량 분포')
plt.show()

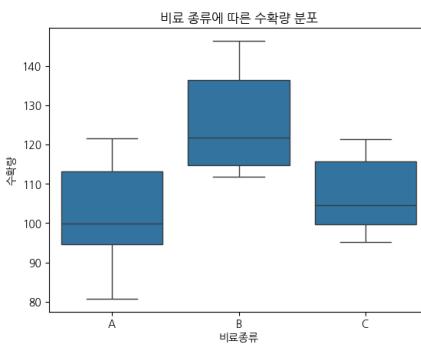
# -----
# 2. 분산분석 (ANOVA)
# H0: 세 비료의 평균 수확량은 모두 같다.
# H1: 적어도 하나 이상의 비료 평균 수확량은 다르다.
# -----
f_statistic, p_value = f_oneway(
    fertilizer_a,
    fertilizer_b,
    fertilizer_c
)
```

```

print(f"ANOVA F-statistic: {f_statistic:.4f}")
print(f"ANOVA p-value: {p_value:.4f}")

# -----
# 3. 사후분석 (Tukey's HSD)
# -----
if p_value < 0.05:
    tukey_result = pairwise_tukeyhsd(
        endog=df['수학량'], # 종속변수
        groups=df['비료종류'], # 그룹 변수
        alpha=0.05
    )
    print("\nTukey's HSD 사후분석 결과:")
    print(tukey_result)
else:
    print("\nANOVA 결과가 유의미하지 않으므로 사후분석을 수행하지 않습니다.")

```



ANOVA F-statistic: 10.3455
ANOVA p-value: 0.0005

Tukey's HSD 사후분석 결과: Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1 group2 meandiff p-adj lower upper reject

A	B	23.6754	0.0006	10.0402	37.3106	True
A	C	4.8429	0.6569	-8.7923	18.4781	False
B	C	-18.8325	0.0054	-32.4677	-5.1972	True

ANOVA 결과 해석

- F-statistic = 10.3455
- p-value = 0.0005 (< 0.05)

해석

일원분산분석 결과, 비료 종류에 따른 평균 수학량에는 통계적으로 유의미한 차이가 존재한다.
즉, 세 비료의 평균 수학량이 모두 같다는 귀무가설은 기각된다.

👉 ANOVA는 “차이가 있다”지만 알려주므로,
어떤 비료 간에 차이가 있는지는 사후분석으로 확인해야 한다.

박스플롯(시각화) 해석

- **비료 B**

- 중앙값이 가장 높고
- 전체 분포도 A, C보다 위쪽에 위치

- **비료 A와 C**

- 중앙값과 분포가 서로 유사
- 일부 겹침이 큼

시각적 해석

시각적으로도 비료 B는

다른 비료에 비해 수확량이 높게 분포하며,

Tukey HSD 사후분석 결과 해석 (핵심)

비교	p-adj	reject	해석
A vs B	0.0006	True	B가 A보다 유의미하게 큼
A vs C	0.6569	False	A와 C 차이 없음
B vs C	0.0054	True	B가 C보다 유의미하게 큼

비료 B는 비료 A 및 C와 비교했을 때

평균 수확량이 통계적으로 유의미하게 높다.

반면, 비료 A와 C 사이에는 유의미한 차이가 없다.

종합 결론 (보고서용 한 문단)

일원분산분석 결과 비료 종류에 따른 평균 수확량에는 유의미한 차이가 있었으며($p < 0.001$),

Tukey의 HSD 사후분석 결과 비료 B는 비료 A 및 C보다 수확량이 유의미하게 높았다.

비료 A와 C 간에는 통계적으로 유의미한 차이가 나타나지 않았다.

따라서 본 실험 조건에서 생산량 증대 효과가 가장 우수한 비료는 **비료 B**로 판단된다.

문제 5-1

수행 과제: ANOVA의 귀무가설, 즉 "세 비료의 효과가 완전히 똑같다면" 어떤 F-통계량 값들이 나타날 수 있는지 시뮬레이션으로 확인해봅시다. 이를 통해 우리가 실제로 관찰한 F-통계량이 얼마나 극단적인 값인지 직접 확인해 보세요.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import f_oneway

# 1. 시뮬레이션 설정
# H0가 사실이라고 가정: 세 그룹의 평균은 모두 같다.
# 모든 데이터를 합쳐서 하나의 거대한 모집단(H0)을 만들고, 그 평균과 표준편차를 구합니다.
combined_harvest = df['수확량']
grand_mean = combined_harvest.mean()
grand_std = combined_harvest.std()
num_simulations = 5000
simulated_f_stats = []

# [작성] for 반복문을 사용하여 시뮬레이션을 5,000번 수행하세요.
# 각 반복마다, H0 모집단에서 10개씩 3개의 가상 그룹을 생성하고,
# 이 가상 그룹들로 ANOVA를 수행하여 F-통계량을 계산하여 simulated_f_stats 리스트에 추가하세요.
```

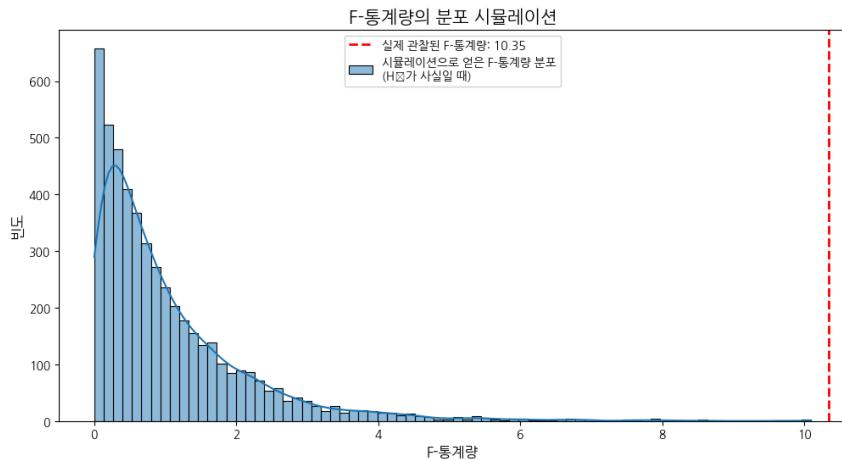
```

for _ in range(num_simulations):
    #  $H_0$ 가 사실이라고 가정: 세 그룹 평균이 동일(= grand_mean)
    sim_a = np.random.normal(loc=grand_mean, scale=grand_std, size=10)
    sim_b = np.random.normal(loc=grand_mean, scale=grand_std, size=10)
    sim_c = np.random.normal(loc=grand_mean, scale=grand_std, size=10)

    # 가상 데이터로 ANOVA 수행 → F 통계량 저장
    f_sim, _ = f_oneway(sim_a, sim_b, sim_c)
    simulated_f_stats.append(f_sim)

# 2. 시각화
plt.figure(figsize=(12, 6))
sns.histplot(simulated_f_stats, kde=True, label='시뮬레이션으로 얻은 F-통계량 분포\n(H0가 사실일 때)')
plt.axvline(x=f_statistic, color='red', linestyle='--', linewidth=2,
            label=f'실제 관찰된 F-통계량: {f_statistic:.2f}')
plt.title("F-통계량의 분포 시뮬레이션", fontsize=15)
plt.xlabel("F-통계량", fontsize=12)
plt.ylabel("빈도", fontsize=12)
plt.legend()
plt.show()

```



1)ANOVA 검정의 p-값으로 어떤 결론을 내릴 수 있나요?

- 관찰된 값
 - $F\text{-통계량} \approx 10.35$
 - $\text{ANOVA p-value} = 0.0005$

해석

ANOVA의 p-값이 유의수준 0.05보다 매우 작으므로,
“세 비료의 평균 수확량이 모두 동일하다”는 귀무가설은 기각된다.

⚠ 중요한 해석 포인트

- 이 결과는 “세 비료가 모두 동일한 효과를 가진다”는 뜻이 아니다.
- 정확한 의미는
 - 👉 “적어도 하나 이상의 비료는 다른 비료와 평균 수확량이 다르다”는 것이다.

즉, 차이가 존재함은 확실하지만, 누가 다른지는 아직 모른다 → 사후분석 필요.

2) Tukey's HSD 결과에서 `reject=True` 인 조합과 그 의미는?

Tukey 결과 요약

비교	reject	의미
A vs B	True	A와 B는 평균 수확량이 유의미하게 다름
A vs C	False	A와 C는 유의미한 차이 없음
B vs C	True	B와 C는 평균 수확량이 유의미하게 다름

meandiff 열의 의미

- `meandiff = group1 - group2`

- **부호:** 어느 쪽 평균이 더 큰지

- **절댓값:** 평균 차이의 크기

예:

- $A - B = +23.68 \rightarrow B$ 가 A보다 약 23.7만큼 더 큼
- $B - C = -18.83 \rightarrow B$ 가 C보다 약 18.8만큼 더 큼

👉 종합하면

비료 B가 A와 C보다 평균 수확량이 유의미하게 높다는 뜻이다.

3) 시뮬레이션 그래프를 포함한 종합 해석 & 비즈니스 추천

📊 시뮬레이션 그래프 해석 (지금 이미지 기준)

- 히스토그램은

👉 귀무가설이 사실일 때(세 비료 효과가 같을 때)

나올 수 있는 F-통계량의 분포

- 분포의 대부분은 **0~3** 근처

- 실제 관찰된 $F \approx 10.35$ (빨간 점선) 는

👉 분포의 극단적인 오른쪽 꼬리에 위치

의미

세 비료의 효과가 정말 같다면

이 정도로 큰 F-통계량은 거의 발생하지 않는다.

따라서 현재 관찰된 결과는 우연으로 보기 어렵다.

이 시각적 판단은 ANOVA의 매우 작은 **p-value(0.0005)** 와 정확히 일치한다.

4) 연구에 대한 비즈니스 관점 최종 추천

✓ 추천 비료: 비료 B

이유 (데이터 + 비즈니스 관점)

통계 분석 및 시뮬레이션 결과, 비료 B는 다른 비료(A, C)에 비해 수확량을 유의미하게 증가시키는 효과를 보였다. 이는 우연이 아닌 통계적으로 검증된 차이이며,

생산량 증가 → 매출 증대 → 농가 경쟁력 강화로 이어질 가능성이 높다.

반면 비료 A와 C는 서로 유의미한 차이가 없으므로, 한정된 연구·마케팅 자원을 고려할 때

비료 B에 집중하는 전략이 가장 효율적이다.

🔑 최종 결론

“세 비료의 효과는 같지 않으며, 그 차이를 만들어낸 것은 비료 B이고, 이 결과는 통계적으로도 시각적으로도 매우 강력하다.”

