

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

## 패키지 가져오기

```

from pandas import read_excel, melt
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm
from matplotlib import pyplot as plt
import seaborn as sb
import sys
import os

sys.path.append(os.path.dirname(os.path.dirname(os.getcwd())))
from helper import normality_test, equal_variance_test, independence_test

```

## 예제 (1)

태아의 머리 둘레 측정 데이터.

3명의 태아를 대상으로 4명의 관측자가 측정.

이를 통해서 초음파로 태아의 머리 둘레측정 데이터가 재현성이 있는지를 조사

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

## 패키지 가져오기

## 예제 (1)

## 1. 데이터 가져오기

## 2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

## 3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

## 데이터 가져오기

## 2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

## 추론적 데이터 분석

## 1. 데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E02/alrtman.xlsx")
df
```

	태아번호	측정자번호	머리둘레
0	1	1	14.3
1	1	1	14.0
2	1	1	14.8
3	1	2	13.6
4	1	2	13.6
5	1	2	13.8
6	1	3	13.9
7	1	3	13.7
8	1	3	13.8
9	1	4	13.8
10	1	4	14.7
11	1	4	13.9
12	2	1	19.7
13	2	1	19.9
14	2	1	19.8

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

	태아번호	측정자번호	머리둘레
15	2	2	19.8
16	2	2	19.3
17	2	2	19.8
18	2	3	19.5
19	2	3	19.8
20	2	3	19.5
21	2	4	19.8
22	2	4	19.6
23	2	4	19.8
24	3	1	13.0
25	3	1	12.6
26	3	1	12.9
27	3	2	12.4
28	3	2	12.8
29	3	2	12.5
30	3	3	12.8
31	3	3	12.7
32	3	3	12.5

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

	태아번호	측정자번호	머리둘레
33	3	4	13.0
34	3	4	12.9
35	3	4	13.8

## 2. 탐색적 데이터 분석

## 측정자에 따른 태아의 머리둘레 확인

그림을 보면 태아 3명의 머리둘레는 차이가 있어보이지만 이것이 관측자와 상호작용이 있는 것인지 분석을 통해 알아보고자 함

```
plt.rcParams["font.family"] = 'AppleGothic' if sys.platform == 'darwin'
plt.rcParams["font.size"] = 12
plt.rcParams["figure.figsize"] = (10, 5)
plt.rcParams["axes.unicode_minus"] = False
```

```
plt.figure()
sb.boxplot(data=df, x='태아번호', y='머리둘레', hue='측정자번호')
plt.show()
plt.close()
```

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

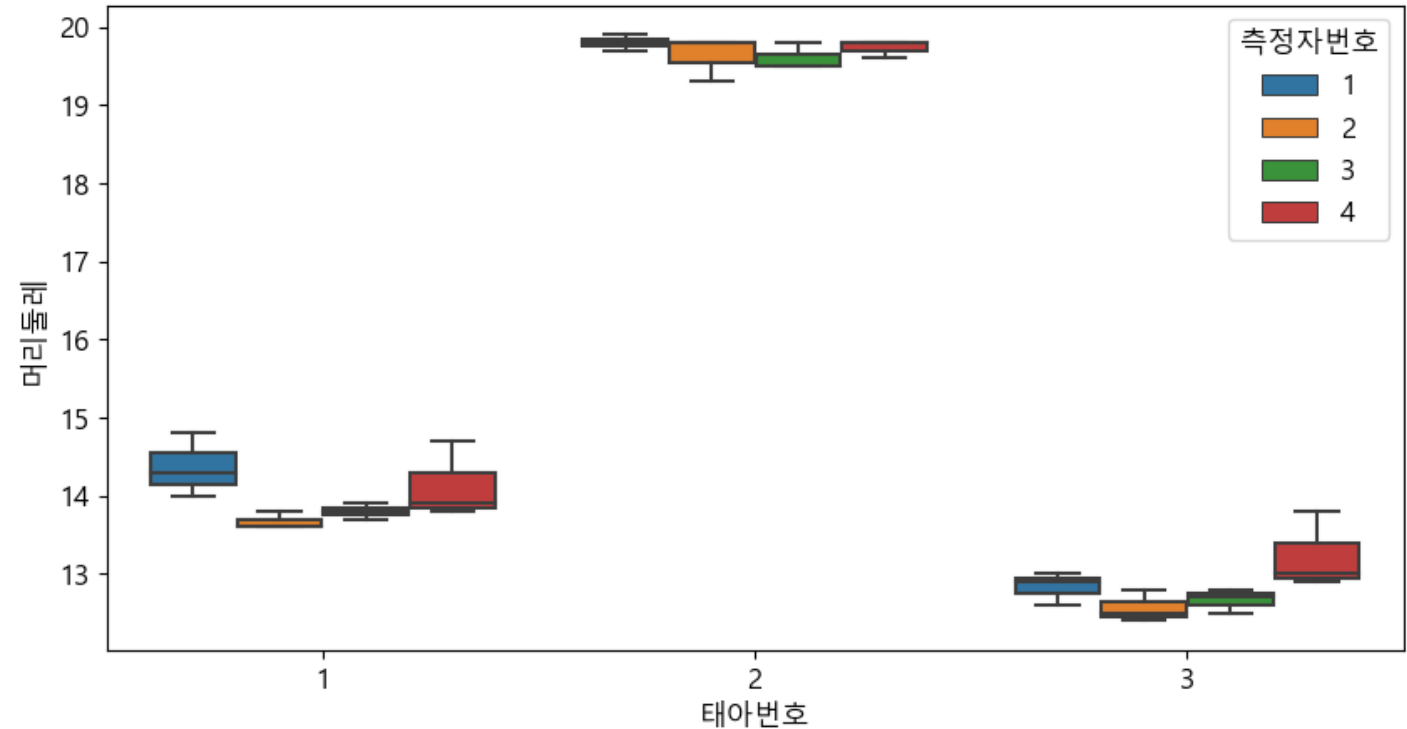
데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석



## 3. 추론적 데이터 분석

## 가설설정

가설	내용
귀무가설	태아의 머리 둘레 차이는 측정자와 연관이 없다.
대립가설	태아의 머리 둘레 차이는 측정자와 연관이 있다.

## 정규성, 등분산성, 독립성 검정

예제 데이터이므로 충족한다고 가정하고 분석을 진행함

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

```
all_test(df['태아번호'], df['측정자번호'], df['머리둘레'])
```

			statistic	p-value	result
condition	test	field			
정규성	shapiro	태아번호	0.794544	1.271364e-05	False
		측정자번호	0.857769	2.887842e-04	False
		머리둘레	0.742801	1.429796e-06	False
	normaltest	태아번호	23.481208	7.963801e-06	False
		측정자번호	14.053919	8.876264e-04	False
		머리둘레	24.573578	4.612277e-06	False
		태아번호 vs 측정자번호	0.250000	2.122867e-01	True
	ks_2samp	측정자번호 vs 머리둘레	1.000000	4.519646e-21	False
		머리둘레 vs 태아번호	1.000000	4.519646e-21	False

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

			statistic	p-value	result
condition	test	field			
등분산성	Bartlett	태아번호 vs 측정자번호 vs 머리둘레	64.853335	8.265656e-15	False
	Fligner	태아번호 vs 측정자번호 vs 머리둘레	12.782139	1.676462e-03	False
	Levene	태아번호 vs 측정자번호 vs 머리둘레	13.063350	8.583289e-06	False
독립성	Chi2	태아번호 vs 측정자번호 vs 머리둘레	33.814972	9.999218e-01	True

## 이원분산분석 수행

scipy.stats 패키지는 이원분산분석을 지원하지 않음

- 머리둘레 와 태아
- 머리둘레 와 측정자
- 머리둘레 와 태아 vs 측정자

세 가지 상황에 따른 평균에 유의미한 차이가 있는지 분석

```
formula = '머리둘레 ~ C(태아번호)+C(측정자번호)+C(태아번호):C(측정자번호)'
lm = ols(formula, df).fit()
anova = anova_lm(lm)
anova
```

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

	df	sum_sq	mean_sq	F	PR(>F)
C(태아번호)	2.0	324.008889	162.004444	2113.101449	1.051039e-27
C(측정자번호)	3.0	1.198611	0.399537	5.211353	6.497055e-03
C(태아번호):C(측정자번호)	6.0	0.562222	0.093704	1.222222	3.295509e-01
Residual	24.0	1.840000	0.076667	NaN	NaN

## 결과해석

p-value가 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택한다.

즉 측정자(측정장비)와 태아의 머리둘레값에는 연관성이 있다고 할 수 있다.

## 예제 (2)

한 식물학자는 식물의 성장이 햇빛 노출과 관수 빈도에 의해 영향을 받는지 여부를 알고 싶어한다.

실험은 30개의 씨앗을 심고 햇빛 노출과 물 주기를 다른 조건으로 하여 두 달 동안 자라게 한 후 각 식물의 높이를 인치 단위로 기록하였다.

## 데이터 가져오기

필드	설명	값
water	각 식물에 물을 주는 빈도	매일: daily, 매주: weekly



## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

## 1. 데이터 가져오기

## 2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

## 3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

## 2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

## 추론적 데이터 분석

필드	설명	값
sun	각 식물이 받은 햇빛 노출량	낮음: low, 중간: med, 높음: high
height	2개월 후 각 식물의 높이	인치

```
df = read_excel("https://data.hossam.kr/E02/plant_growth.xlsx")
df
```

	water	sun	height
0	daily	low	6
1	daily	low	6
2	daily	low	6
3	daily	low	5
4	daily	low	6
5	daily	med	5
6	daily	med	5
7	daily	med	6
8	daily	med	4
9	daily	med	5
10	daily	high	6
11	daily	high	6

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

	water	sun	height
12	daily	high	7
13	daily	high	8
14	daily	high	7
15	weekly	low	3
16	weekly	low	4
17	weekly	low	4
18	weekly	low	4
19	weekly	low	5
20	weekly	med	4
21	weekly	med	4
22	weekly	med	4
23	weekly	med	4
24	weekly	med	4
25	weekly	high	5
26	weekly	high	6
27	weekly	high	6
28	weekly	high	7
29	weekly	high	8

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

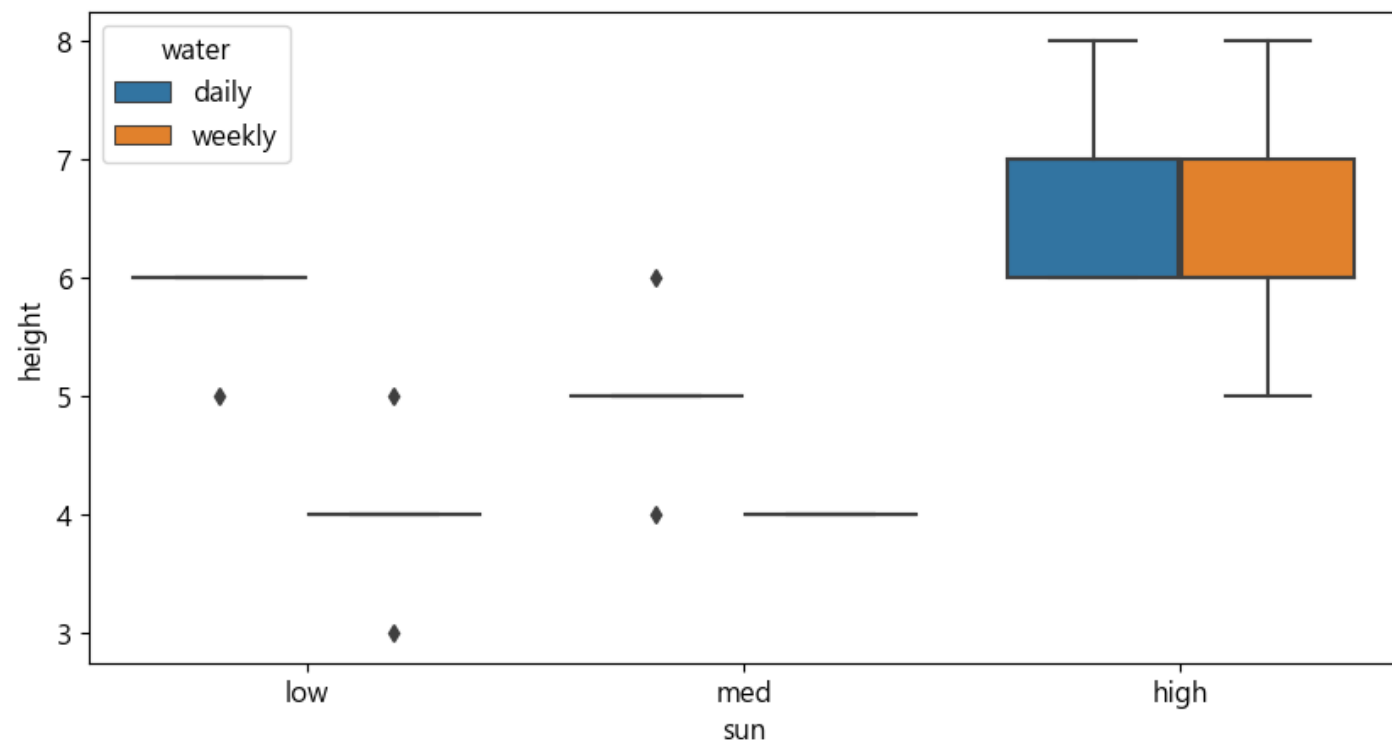
데이터 전처리

추론적 데이터 분석

## 2. 탐색적 데이터 분석

요인에 따른 식물의 높이 확인

```
plt.figure()
sb.boxplot(data=df, x='sun', y='height', hue='water')
plt.show()
plt.close()
```



## 데이터 전처리

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

## 1. 데이터 가져오기

## 2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

## 3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

## 2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

## 추론적 데이터 분석

```
df2 = df.copy()
df2['sun'] = df2['sun'].map({'low': 0, 'med': 1, 'high': 2})
df2['water'] = df2['water'].map({'daily': 0, 'weekly': 1})
df2
```

	water	sun	height
0	0	0	6
1	0	0	6
2	0	0	6
3	0	0	5
4	0	0	6
5	0	1	5
6	0	1	5
7	0	1	6
8	0	1	4
9	0	1	5
10	0	2	6
11	0	2	6
12	0	2	7
13	0	2	8

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

	water	sun	height
14	0	2	7
15	1	0	3
16	1	0	4
17	1	0	4
18	1	0	4
19	1	0	5
20	1	1	4
21	1	1	4
22	1	1	4
23	1	1	4
24	1	1	4
25	1	2	5
26	1	2	6
27	1	2	6
28	1	2	7
29	1	2	8

## 추론적 데이터 분석

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

## 정규성, 등분산성, 독립성 확인

```
all_test(df2['sun'], df2['water'], df2['height'])
```

			statistic	p-value	result
condition	test	field			
정규성	shapiro	sun	0.795503	5.434527e-05	False
		water	0.638237	2.211095e-07	False
		height	0.917132	2.261243e-02	False
	normaltest	sun	17.211505	1.830498e-04	False
		water	279.252501	2.296624e-61	False
		height	1.140421	5.654064e-01	True
	ks_2samp	sun vs water	0.333333	7.088799e-02	True
		water vs height	1.000000	1.691123e-17	False
		height vs sun	1.000000	1.691123e-17	False
등분산성	Bartlett	sun vs water vs height	23.058466	9.838245e-06	False
	Fligner	sun vs water vs height	22.593788	1.241141e-05	False
	Levene	sun vs water vs height	9.023098	2.748167e-04	False
독립성	Chi2	sun vs water vs height	32.576490	9.971957e-01	True

## 이원분산분석 수행

## Two-way ANOVA (이원분산분석)

## #01. 작업 준비

패키지 가져오기

## 예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

측정자에 따른 태아의 머리  
둘레 확인

3. 추론적 데이터 분석

가설설정

정규성, 등분산성, 독립성  
검정

이원분산분석 수행

결과해석

## 예제 (2)

데이터 가져오기

2. 탐색적 데이터 분석

요인에 따른 식물의 높이  
확인

데이터 전처리

추론적 데이터 분석

```
model = ols('height ~ C(water) + C(sun) + C(water):C(sun)', data=df).fit()
anova_lm(model)
```

	df	sum_sq	mean_sq	F	PR(>F)
<b>C(water)</b>	1.0	8.533333	8.533333	16.0000	0.000527
<b>C(sun)</b>	2.0	24.866667	12.433333	23.3125	0.000002
<b>C(water):C(sun)</b>	2.0	2.466667	1.233333	2.3125	0.120667
<b>Residual</b>	24.0	12.800000	0.533333	NaN	NaN

## 결과해석

물과 태양에 대한 p-값이 모두 **0.05** 미만이므로 두 요인 모두 식물 높이에 통계적으로 유의미한 영향을 미친다.

그리고 교호작용 효과(.120667)에 대한 p-값이 **0.05** 이상이므로 햇빛 노출과 물 주기 사이에 유의한 교호작용 효과가 없음을 알 수 있다.

## 교호작용(Interaction)

한 요인의 효과가 다른 요인의 수준에 의존하는 경우