

03

## 이원 배치 분산분석 실습

[이원 배치 분산 분석 실습]

00:11

# fomula 안에 수식을 작성하는 방법을 기억해주세요~

**3.6 Python을 활용한 이원배치 분산분석**

참수 사용법 Python으로 양방향 ANOVA를 수행하는 훨씬 더 쉬운 방법이 있습니다. 많은 R- 패키지 (예 : lm)와 유사한 모델 표기법을 가진 Statsmodel을 사용할 수 있습니다. 모델의 공식화부터 시작합니다.

- formula = 'len ~ C(supp) + C(dose) + C(supp):C(dose)'
- model = ols(formula, data).fit()
- aov\_table = anova\_lm(model, type=2)

**참수 사용 예제**

Python에 내장된 mtcars 데이터는 32개의 차종에 대한 다양한 특성과 단위 연료당 주행거리를 담고 있다. am변수는 변속기 종류이며, cyl변수는 실린더 개수를 의미한다. 데이터를 분석에 적절한 형태로 처리한 후, 변속기 종류(am변수)와 실린더 개수(cyl)에 따라 주행거리(mpg변수) 평균에 유의미한 차이가 존재하는지 이원 분산분석을 수행하고, 그 결과를 해석해보자

- 검정을 수행하기에 앞서 설정할 수 있는 가설은 아래와 같다.

- 가. 상호작용효과 검정에 대한 가설
  - 1) 귀무가설 H0: 변속기 종류와 실린더 개수 간에는 상호작용 효과가 없다.
  - 2) 대립가설 H1: 변속기 종류와 실린더 개수 간에는 상호작용 효과가 있다.
- 나. 주효과 검정에 대한 가설
  - 1) 귀무가설 H0: 실린더 개수에 따른 주행거리의 차이는 존재하지 않는다.
  - 2) 대립가설 H1: 실린더 개수에 따른 주행거리의 차이는 존재한다.
  - 3) 귀무가설 H0: 변속기 종류에 따른 주행거리의 차이는 존재하지 않는다.
  - 4) 대립가설 H1: 변속기 종류에 따른 주행거리의 차이는 존재한다.

[mtcars EDA 및 전처리]

02:12

In [10]:

```
import pandas as pd
import numpy as np
%matplotlib inline
```

In [2]:

```
# 데이터 불러오기
mtcars = pd.read_csv('mtcars.csv')
```

In [3]:

```
mtcars
```

Out[3]:

	Unnamed: 0	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	Valiant	18.1	6	225.0	105	2.78	3.460	20.22	1	0	3	1
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
7	Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
8	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
9	Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4

[분석 시행 및 결과해석]

# 그래프가 교차할 때는 검정 할 필요 없이 두 개의 변수 간에 교호작용이 있다고 할 수 있지만 교차점이 없을 때에는 이원분산분석 결과표에 나타난 p-value값으로 해석하는 것이 맞습니다.

04:09

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

```

In [7]: ## 분산분석 수행
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm

formula = 'mpg ~ C(cyl)+C(am)'
model = ols(formula, mtcars).fit()
aov_table = anova_lm(model, typ=2)

In [8]: aov_table
Out[8]:

```

	sum_sq	df	F	PR(>F)
C(cyl)	456.409921	2.0	24.819011	9.354735e-07
C(am)	36.766919	1.0	3.998759	5.608373e-02
C(cyl):C(am)	25.436511	2.0	1.383233	2.696140e-01
Residual	239.059167	26.0	NaN	NaN

- cyl변수에 대한 p-value는 0.05보다 작으므로, 주효과 검정에서 1)귀무가설을 기각한다. 따라서 실린더 개수에 따라 주행거리 간 유의미한 차이는 존재한다고 해석할 수 있다.
- am변수에 대한 p-value는 0.0561로 0.05보다 크므로, 주효과 검정에서 '변속기 종류에 따른 주행거리 평균 간 차이는 존재하지 않는다'는 귀무가설을 기각하지 않는다.
- cyl변수와 am변수 간의 상호작용 효과에 대한 검정결과, p-value는 0.2696으로 0.05보다 크므로 귀무가설을 기각하지 않는다. 따라서 실린더 개수와 변속기 종류 간에는 교호작용이 존재하지 않는다는 것을 알 수 있다.