

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

연습문제 3 폴이

작업준비

패키지 가져오기

```
from pandas import read_excel, concat
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
```

문제 1

데이터 가져오기

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

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

df

	독서방법	점수
0	B	6
1	B	9
2	B	2
3	B	16
4	D	7
5	D	7
6	D	12
7	D	10
8	D	16
9	S	11
10	S	7
11	S	4
12	S	7
13	S	7

데이터 전처리 (라벨링)

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

```
df['독서방법'] = df['독서방법'].map({'B': 1, 'D': 2, 'S': 3})  
df
```

	독서방법	점수
0	1	6
1	1	9
2	1	2
3	1	16
4	2	7
5	2	7
6	2	12
7	2	10
8	2	16
9	3	11
10	3	7
11	3	4
12	3	7
13	3	7

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

```
formula = '점수 ~ C(독서방법)'
lm = ols(formula, df).fit()
anova = anova_lm(lm)
anova
```

	df	sum_sq	mean_sq	F	PR(>F)
C(독서방법)	2.0	26.464286	13.232143	0.779403	0.48244
Residual	11.0	186.750000	16.977273	NaN	NaN

문제 2

데이터 가져오기

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

	효소함량	모종성장
0	0	10.8
1	0	9.1
2	0	13.5
3	0	9.2
4	1000	11.1

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	효소함량	모종성장
5	1000	11.2
6	1000	8.2
7	1000	11.3
8	5000	5.4
9	5000	4.6
10	5000	7.4
11	5000	5.0
12	10000	5.8
13	10000	5.3
14	10000	3.2
15	10000	7.5

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  --
0   효소함량    16 non-null     int64
1   모종성장    16 non-null     float64
```

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

```
dtypes: float64(1), int64(1)
memory usage: 388.0 bytes
```

```
df['효소함량'] = df['효소함량'].map({0: 1, 1000: 2, 5000: 3, 10000: 4})
df
```

	효소함량	모종성장
0	1	10.8
1	1	9.1
2	1	13.5
3	1	9.2
4	2	11.1
5	2	11.2
6	2	8.2
7	2	11.3
8	3	5.4
9	3	4.6
10	3	7.4
11	3	5.0
12	4	5.8

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	효소함량	모종성장
13	4	5.3
14	4	3.2
15	4	7.5

```
formula = '모종성장 ~ C(효소함량)'
lm = ols(formula, df).fit()
anova = anova_lm(lm)
anova
```

	df	sum_sq	mean_sq	F	PR(>F)
C(효소함량)	3.0	101.1275	33.709167	12.085748	0.000615
Residual	12.0	33.4700	2.789167	NaN	NaN

문제 3

데이터 가져오기

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

	교육방법	국어	영어
0	1	65	82

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	국어	영어
1	1	87	79
2	1	73	85
3	1	79	60
4	1	81	65
5	1	69	70
6	1	55	79
7	1	76	80
8	1	77	76
9	1	70	90
10	1	88	68
11	2	75	72
12	2	69	75
13	2	83	93
14	2	81	85
15	2	72	60
16	2	79	65
17	2	85	88
18	2	89	92

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	국어	영어
19	2	90	83
20	2	77	93
21	2	95	78

```
df1 = df.filter(['교육방법', '국어'])
df1.rename(columns={'국어': '점수'}, inplace=True)
df1['과목'] = '국어'
df1
```

	교육방법	점수	과목
0	1	65	국어
1	1	87	국어
2	1	73	국어
3	1	79	국어
4	1	81	국어
5	1	69	국어
6	1	55	국어
7	1	76	국어
8	1	77	국어
9	1	70	국어

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
10	1	88	국어
11	2	75	국어
12	2	69	국어
13	2	83	국어
14	2	81	국어
15	2	72	국어
16	2	79	국어
17	2	85	국어
18	2	89	국어
19	2	90	국어
20	2	77	국어
21	2	95	국어

```
df2 = df.filter(['교육방법', '영어'])
df2.rename(columns={'영어': '점수'}, inplace=True)
df2['과목'] = '영어'
df2
```

	교육방법	점수	과목
0	1	82	영어

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
1	1	79	영어
2	1	85	영어
3	1	60	영어
4	1	65	영어
5	1	70	영어
6	1	79	영어
7	1	80	영어
8	1	76	영어
9	1	90	영어
10	1	68	영어
11	2	72	영어
12	2	75	영어
13	2	93	영어
14	2	85	영어
15	2	60	영어
16	2	65	영어
17	2	88	영어
18	2	92	영어

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
19	2	83	영어
20	2	93	영어
21	2	78	영어

```
df3 = concat([df1, df2])
df3
```

	교육방법	점수	과목
0	1	65	국어
1	1	87	국어
2	1	73	국어
3	1	79	국어
4	1	81	국어
5	1	69	국어
6	1	55	국어
7	1	76	국어
8	1	77	국어
9	1	70	국어
10	1	88	국어

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
11	2	75	국어
12	2	69	국어
13	2	83	국어
14	2	81	국어
15	2	72	국어
16	2	79	국어
17	2	85	국어
18	2	89	국어
19	2	90	국어
20	2	77	국어
21	2	95	국어
0	1	82	영어
1	1	79	영어
2	1	85	영어
3	1	60	영어
4	1	65	영어
5	1	70	영어
6	1	79	영어

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
7	1	80	영어
8	1	76	영어
9	1	90	영어
10	1	68	영어
11	2	72	영어
12	2	75	영어
13	2	93	영어
14	2	85	영어
15	2	60	영어
16	2	65	영어
17	2	88	영어
18	2	92	영어
19	2	83	영어
20	2	93	영어
21	2	78	영어

```
df3['과목'] = df3['과목'].map({'국어': 1, '영어': 2})
df3
```

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
0	1	65	1
1	1	87	1
2	1	73	1
3	1	79	1
4	1	81	1
5	1	69	1
6	1	55	1
7	1	76	1
8	1	77	1
9	1	70	1
10	1	88	1
11	2	75	1
12	2	69	1
13	2	83	1
14	2	81	1
15	2	72	1
16	2	79	1
17	2	85	1

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
18	2	89	1
19	2	90	1
20	2	77	1
21	2	95	1
0	1	82	2
1	1	79	2
2	1	85	2
3	1	60	2
4	1	65	2
5	1	70	2
6	1	79	2
7	1	80	2
8	1	76	2
9	1	90	2
10	1	68	2
11	2	72	2
12	2	75	2
13	2	93	2

연습문제 3 폴이

작업준비

패키지 가져오기

문제 1

데이터 가져오기

데이터 전처리 (라벨링)

문제 2

데이터 가져오기

문제 3

데이터 가져오기

	교육방법	점수	과목
14	2	85	2
15	2	60	2
16	2	65	2
17	2	88	2
18	2	92	2
19	2	83	2
20	2	93	2
21	2	78	2

```
formula = '점수 ~ C(교육방법)*C(과목)'
lm = ols(formula, df3).fit()
anova = anova_lm(lm)
anova
```

	df	sum_sq	mean_sq	F	PR(>F)
C(교육방법)	1.0	355.113636	355.113636	3.845870	0.056857
C(과목)	1.0	0.204545	0.204545	0.002215	0.962695
C(교육방법):C(과목)	1.0	14.204545	14.204545	0.153835	0.696979
Residual	40.0	3693.454545	92.336364	NaN	NaN