

7장 연습문제

모든 답에는 계산을 위한 코딩 과정이 함께 포함되어야 합니다.

학문적 성취보다 학문적 양심을 지키는 것이 훨씬 중요합니다.

과제 데드라인이 임박했다는 이유로 copy를 해서 제출하지 않도록 당부드립니다.

1. 전국 고등학생들 중 남학생과 여학생을 각각 150명씩 선별하여 수학 과목에 대한 선호도 설문조사를 시행하였다. 그 결과 수학 과목을 선호한다고 응답한 학생이 남학생과 여학생의 순서대로 54, 43명이었다. 남학생과 여학생이 수학을 선호하는 비율을 순서대로 p_1, p_2 라 하자. (총 20점)

(1) p_1 이 p_2 보다 높다는 유의한 증거가 있는가? 유의수준 5% 이내에서 검정하라. (10점)

(2) $p_1 - p_2$ 의 95% (근사) 신뢰구간을 구하여라. (10점)

-
1. data1.csv는 900명의 4년제 대학생을 대상으로 학년과 안경 착용 유무를 조사한 것이다. 다음과 같은 형태를 가진다.

In []:

```
data1=pd.read_csv('~/data1.csv')
data1.head()
```

Out []:

	Unnamed: 0	학년	안경 유무
0	0	2	0
1	1	4	0
2	2	1	1
3	3	1	1
4	4	4	0

안경을 쓴 경우 1, 쓰지 않은 경우 0의 값을 가지도록 기재되었다. 위 데이터를 이용하여 유의수준 5% 이내에서 학년과 안경 착용 사이에 연관이 있는지 검정하려 한다. (총 30점)

- (1) 검정하고자 하는 가설을 밝히시오. (5점)
(2) 검정을 위해, 학년을 행으로, 안경 착용 여부를 열로 하는 분할표(관측도수 행렬)를 구하시오. (15점)
(3) 검정통계량의 관측값, p-value를 계산하고, 검정 결과의 의미를 쓰시오. (10점)
-