

## 통계학실험 9장 추가과제

예제4 서로 다른 네종류의 비료의 효과를 비교하기 위해 세 지역에 같은 작물을 심고 수확량을 비교하였다. 수확량은 다음과 같다.

변수명	region1	region2	region3
fertilizer1	12.1	12.8	13.9
fertilizer2	13.6	18.4	17.5
fertilizer3	14.2	17.2	19.2
fertilizer4	21.7	21.4	17.5

비료에 따라 수확량이 달라진다고 할 수 있는가? 적절한 가설을 쓰고 유의수준 5%에서 이를 검정하시오.

### 예제5

mlb\_players\_18는 2018 시즌 미국 메이저리그에서 100타석 이상의 기록이 있는 타자 429명의 통계를 정리한 자료이다. mlb\_players\_18 자료는 다음의 변수들로 이루어져 있다.

(첨부된 mlb\_players\_18.csv 파일 이용하세요.)

- name: 선수 이름
- team: 소속 팀
- position: 선수의 수비 위치 (C, 1B, 2B, 3B, LF, CF, RF, SS)
- AB: 타석 수
- H: 총 안타 수 (아래 doubles, triples, HR 개수를 모두 포괄한 수치)
- doubles: 총 2루타 수
- triples: 총 3루타 수
- HR: 총 홈런 수
- AVG: 타율
- OBP: 출루율

이 문제에서는 선수의 수비 위치에 따라 타율이 달라지는지 검정하고자 한다.

5-1 선수의 수비 위치에 따른 타율을 상자그림으로 표현해보자. 수비 위치에 따라 타율이 다르다고 예상할 수 있는가?

5-2 수비 위치가 같은 선수들의 타율이 정규분포를 따르는지 분위수대조도를 이용해 설명하시오. 다음 코드를 이용하면 한 화면에 9개의 그래프를 그릴 수 있다.

```
> par(mfrow = c(3,3))
```

1	2	3
4	5	6
7	8	9

또한 layout(1)을 실행해서 하나의 그래프를 그리도록 다시 변경할 수 있다.

5-3 적절한 가설을 쓰고 유의수준 5%에서 이를 검정하시오

#### 문제 6

MASS::Boston은 Boston의 506개 지역의 주택 가격에 대한 데이터이다. Boston 데이터는 다음의 변수를 포함하고 있다.

- medv : 주택 가격의 중앙값 (단위: \$1000)
- age : 1940년 이전에 지어진 주택의 비율 (단위: %)
- chas : Chales River에 대한 더미 변수 (접해 있을 때 1, 그 외에 0)

이 문제에서는 chas와 age가 가격 medv에 영향을 미치는지 확인하고자 한다.

6-1 age 변수를 세 그룹으로 나누어 Boston\$age\_group이라는 변수로 저장하자. 비율이 50% 이하이면 low, 50% 초과 90% 이하이면 medium, 90% 초과이면 high로 저장하시오.

6-2 새로 만든 age\_group과 chas의 교호작용의 유무를 평균그림을 통해 확인해보시오.

(단, 평균그림의 제목은 Interaction Plot, x축의 이름은 age\_group, y축의 이름은 mean of medv, 범례(legend)의 이름(trace.label)은 chas로 하자. : main = 'Interaction Plot', xlab = 'age\_group', ylab = 'mean of medv', trace.label = 'chas')

6-3 두 변수 각각의 효과와 두 변수의 교호작용의 유무에 대해 적절한 가설을 쓰고 유의수준 5%에서 이를 검정하시오.