2020학년도 ( 1 )학기 과제물(온라인제출용)

**교과목명 : 통계학개론**

**학 번 : 201835-368393**

**성 명 : 여준영**

**연 락 처 : 010-5196-1122**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

o 과제유형 : ( ) 형

o 과 제 명 : 1학기 통계학개론 출석수업대체시험 실험실습 과제

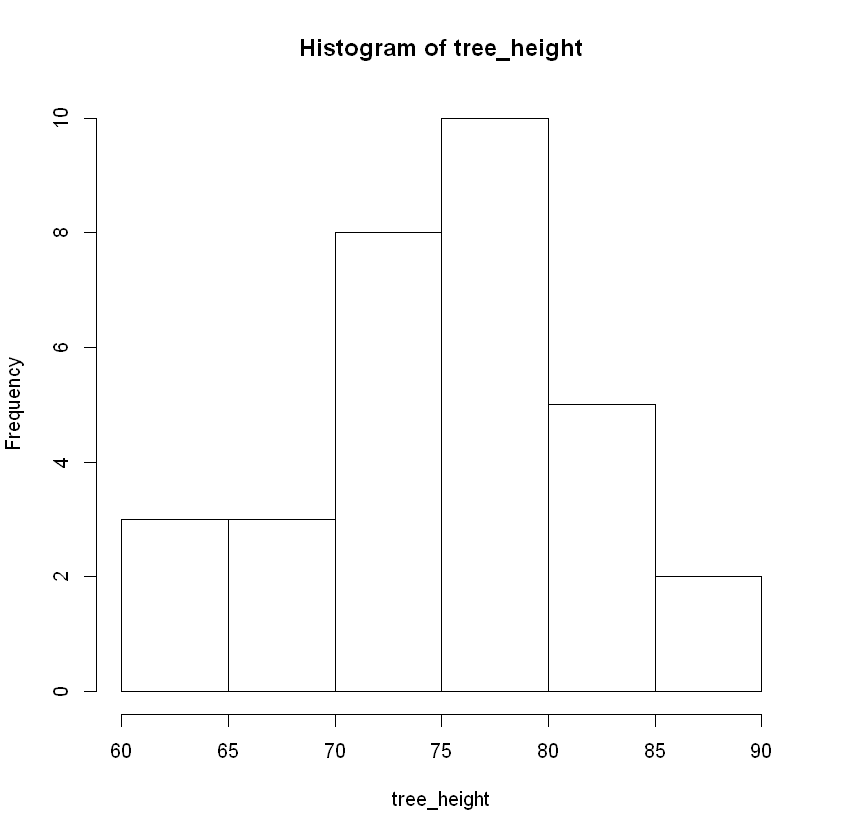
- 이하 과제 작성

※ A4용지 편집 사용

1. (1) 히스토그램

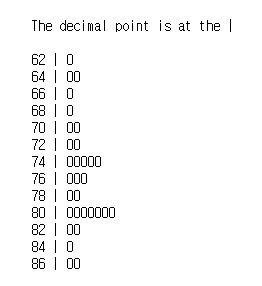
tree\_height <- c(70, 65, 63, 72, 81, 83, 66, 75, 80, 75, 79, 76, 76, 69, 75, 74, 85, 86, 71, 64, 78, 80, 74, 72, 77, 81, 82, 80, 80, 80, 87)

histo <- hist(tree\_height)



(2) 줄기-잎 그림

stem <- stem(tree\_height)



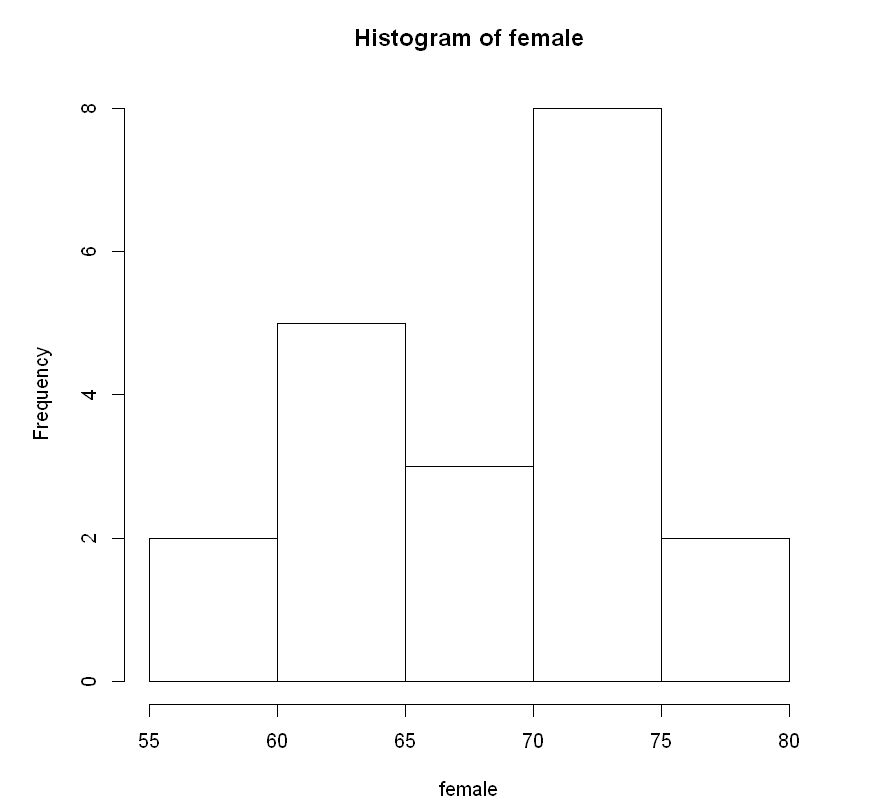
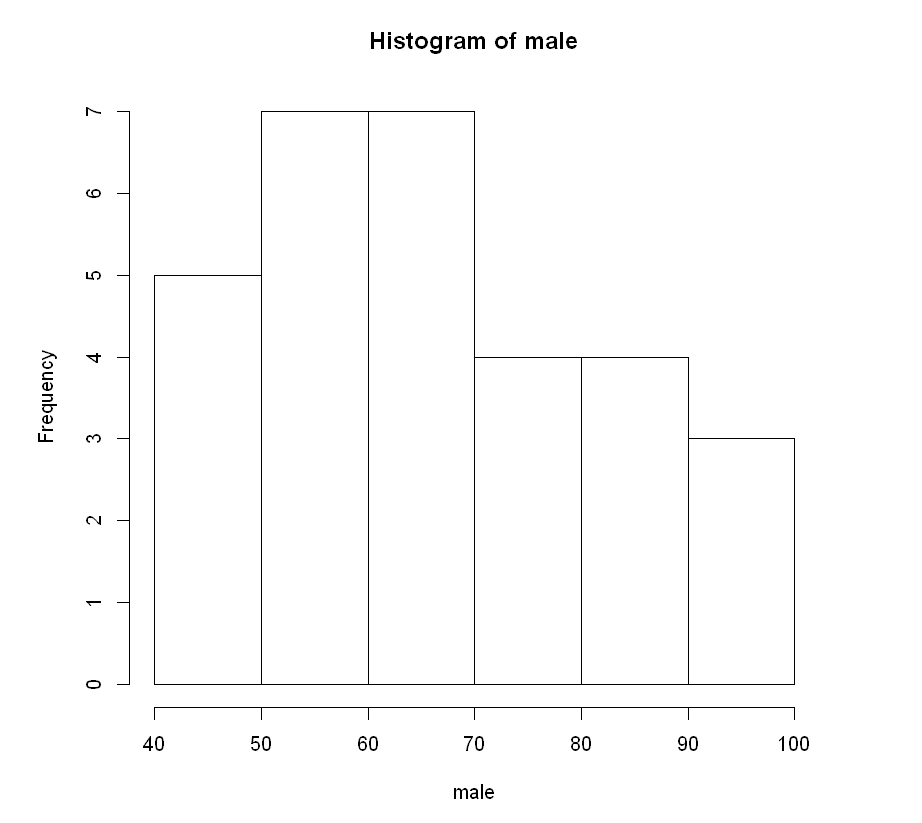
(3) 특징

전체 분포개형은 대략 70~80을 기준으로 좌우대칭성을 띠며, 전체 데이터에서 벗어나는 특이점은 없다.

1. (1) 히스토그램

male <- c(49, 58, 62,86, 82, 72,40, 52, 65,45, 56, 60,48, 50, 64,93, 85, 70,97, 80, 78,58, 60, 67,58, 62, 69,98, 80, 88)

female <- c(60, 68,72, 72,66, 67,65, 61,75, 62,78, 72,62, 79,64, 71,74, 74,58,73)



(2) 변이계수

두 개 이상의 데이터에 대한 퍼짐 정도를 비교하기 위해 두 데이터의 표준편차를 구하여, 비교하는 것은 측정단위가 서로 다르거나 데이터 값의 차이가 커서 무의미한 경우가 많다. 이러한 경우에 사용하는 측도가 **표준편차를 평균으로 나눈 변이계수이다.**

(3)

a. 평균

mean\_male <- mean(male)

>67.7333333333333

mean\_female <- mean(female)

>68.65

b. 표준편차

sd\_male <- sd(male)

>15.8721906736786

sd\_female <- sd(female)

>6.200806399171

c. 변이계수

male.cv = sd\_male / mean\_male

>0.23433352372557

female.cv = sd\_female / mean\_female

>0.0903249293397086

d. 비교

남녀별 입사시험의 성적을 비교함에 있어서 그 데이터의 수가 남자가 여자보다 50%이상 많기 때문에 표준편차만으로는 퍼짐 정도를 비교하기 애매한 상황이다. 따라서 변이계수를 구하여 남자(0.2)가 여자(0.09)보다 퍼짐 정도를 더 크다는 것을 정확하게 측량할 수 있다.



* 분할표

|  |  |  |  |
| --- | --- | --- | --- |
|  | 수학 수강 | 수학 미수강 | 계 |
| 진급 | **4** | **36** | **40** |
| 편입 | **20** | **60** | **80** |
| 계 | **24** | **96** | **120** |

1. 편입생일 확률은 P(편입), 수학을 수강할 확률을 P(수학) 이라고 하자. 조건부 확률의 법칙에 따라서 P(수학|편입) = P(수학∩편입)/P(편입) 이 되므로, (20/120) / (80/120) = 20/80 = 0.25 이다.
2. 진급생일 확률을 P(진급) 이라 했을 때, 문제에 대한 조건부 확률식을 세워보면, P(진급|수학) = P(진급∩수학)/P(수학) 이 되므로, (4/120 / (24/120) = 4/26 ≒ 0.17 이 된다.

(1) 모집단이 정규분포를 따른 다는 것은 표본분포의 크기과 상관없이 정규분포를 따른게 되고, 표본의 크기와 상관없이 표본으로부터 계산되는 통계량과 모수 사이의 관계를 규명할 수 있어 모수의 추정과 검정을 가능하게 하기 때문임

(2)

a. 신뢰구간을 구하기 위해 우선 해당 데이터의 변수화

test <- c(68, 70, 70, 71, 69, 74, 71, 72, 70, 73)mean <- mean(test)

b. 95% 신뢰구간 측정

t.test(test, conf.level = 0.95

One Sample t-test

data: test

t = 123.45, df = 9, p-value = 7.625e-16

alternative hypothesis: true mean is not equal to 0

**95 percent confidence interval:**

**69.50268 72.09732**

sample estimates:

mean of x

70.8

95%의 경우 69.50 <= Z <= 72.09 사이의 신뢰구간을 형성한다.

C. 99% 신뢰구간 측정

t.test(test, conf.level = 0.99)

One Sample t-test

data: test

t = 123.45, df = 9, p-value = 7.625e-16

alternative hypothesis: true mean is not equal to 0

**99 percent confidence interval:**

**68.93626 72.66374**

sample estimates:

mean of x

70.8

99%의 경우 68.96 <= Z <= 72.66 사이에서 신뢰구간을 형성한다.

당연한 결과이지만 99%의 신뢰구간이 95%의 신뢰구간보다 더 넓은 신뢰구간을 보여주고 있다.

1. (1) 600 / 1000 = 0.6  
   (2) [0.6 – 1.96, 0.6 +1.96] = (0.569, 0.630)  
   (3) 우리나라 성인 전체에서 새로운 교육정책에 찬성하는 비율의 신뢰구간은 95% 하에서 0.569 이상 0.630 이하안에서 형성된다.