출석수업 과제물(평가결과물) 표지(온라인제출용)

**교과목명 : 통계학개론**

**학 번 : 202135-368864**

**성 명 : 홍 원 표**

**강 의 실 : 경기(성남) 지역대학 호**

**연 락 처 : 010-5343-4341**

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

- 이하 과제 작성

1. 교재 p.133 예제 5-3의 데이터(어느 대학교 학생들 중 16명의 통계학

시험점수)에 대한 히스토그램과 상자그림을 그리고, 다섯수치요약을

산출하시오. (5점)

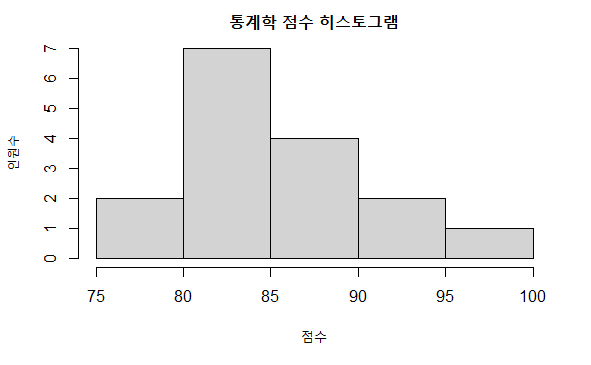
> # 점수 데이터 입력

> *score* = c(88, 83, 83, 85, 94, 88, 91, 96,

+           89, 83, 81, 80, 84, 89, 83, 79)

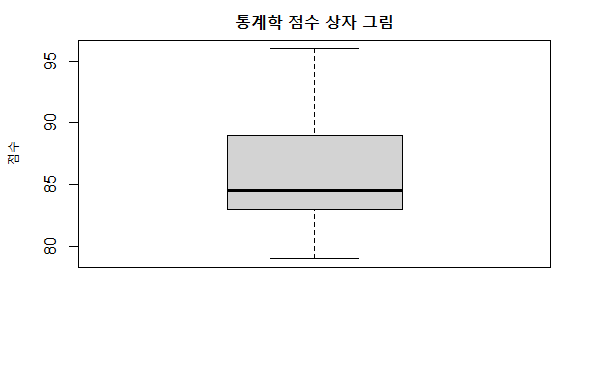
> # 히스토그램

> hist(score, *breaks*=4, *main*="통계학 점수 히스토그램", *xlab*="점수", *ylab*="인원수")



> # 상자 그림

> boxplot(score,  *main*="통계학 점수 상자 그림", *ylab*="점수")



> # 다섯 수치 요약

> fivenum(score)

[1] 79.0 83.0 84.5 89.0 96.0

2. 교재 p.39 예제 2-11의 데이터(학생 10명이 1분당 할 수 있는

윗몸일으키기 개수)를 이용하여, 이 학생들이 대표하는 모집단의

윗몸일으키기 개수의 평균이 25와 같다고 할 수 있는지 유의수준 5%에서

검정하시오. (5점)

윗몸 일으키기 개수의 평균이 25와 같지 않다.

윗몸 일으키기 개수의 평균이 25와 같다.

> # 예제 2-11 데이터 입력

> *counts*=c(25,41,35,8,52,23,32,37,42,28)

# 가설이 ‘같지 않다’이기 때문에 alternative="two.sided" 지정

> t.test(counts, *mu*=25, *alternative*="two.sided")

        One Sample t-test

data:  counts

*t* = 1.8946, *df* = 9, p-*value* = 0.09066

alternative hypothesis: true mean is not equal to 25

95 percent confidence interval:

 23.58392 41.01608

sample estimates:

mean of x

     32.3

p-*value* = **0.09066** 가 유의 수준 5%(**0.05)**보다 높기 때문에 귀무가설을 기각하지 못한다.

**윗몸 일으키기 개수의 평균이 25와 같지 않다.**

3. 교재 p.223 연습문제 1번의 데이터를 이용하여, 근력강화 운동이 체중

감량 효과가 있는지 유의수준 5%에서 검정하시오. (5점)

근력강화 운동이 체중 감량 효과가 없다.

근력강화 운동이 체중 감량 효과가 있다.

> # p.223 문제 1의 데이터 입력

> *before* = c(80, 56, 49, 82, 70)

> *after*  = c(76, 55, 52, 79, 72)

> # before > after 조건이고 두 모평균을 비교하기

> t.test(before, after, *alternative*="greater", *paired*=T)

        Paired t-test

data:  before and after

*t* = 0.43994, *df* = 4, p-*value* = 0.3414

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

 -2.307451       Inf

sample estimates:

mean of the differences

                    0.6

p-*value* = **0.3414** 가 유의 수준 5%(**0.05)**보다 높기 때문에 귀무가설을 기각하지 못한다.

**근력강화 운동이 체중 감량 효과가 없다.**

4. 다음은 세 가지 직업군에 속한 사람들의 월급 데이터이다. 직업군에 따라

월급의 평균이 다르다고 할 수 있는지 일원배치 분산분석으로 검정하시오. (유의수준 5%를 사용한다.) (5점)

|  |  |
| --- | --- |
| 직업군 | 월급 |
| A | 269, 196, 254, 226, 215, 228, 251, 217, 260, 240 |
| B | 320, 281, 336, 303, 294, 354, 315, 259 |
| C | 283, 268, 357, 325, 288, 272, 245, 275, 245, 275, 246, 341 |

직업군에 따라 월급의 평균이 같다.

직업군에 따라 월급의 평균이 다르다.

> # 데이터를 그룹별로 입력

> *A* = c(269, 196, 254, 226, 215, 228, 251, 217, 260, 240)

> *B* = c(320, 281, 336, 303, 294, 354, 315, 259)

> *C* = c(283, 268, 357, 325, 288, 272, 245, 275, 245, 275, 246, 341)

> # 3개의 그룹 데이터를 하나로 만든다.

> X <-c(A, B, C)

> # 명목적 데이터를 생성한다.

> group<-factor(c(rep(1,length(A)), rep(2,length(B)), rep(3,length(C))))

> # formula형식을 위한 데이터 프레임을 생성한다.

> aovdat<-data.frame(X, group)

> # 일원분산분석을 이용하여 검정 실시

> aovmodel<-aov(X~F, *data*=aovdat)

> summary(aovmodel)

            Df Sum Sq Mean Sq F value   Pr(>F)

F            2  25299   12650   12.86 0.000119 \*\*\*

Residuals   27  26550     983

---

Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

p-*value* = 0.000119의 값이 유의 수준 5%(**0.05)** 보다 작기 때문에 귀무가설을 기각하고 대립가설을 채택하게 된다.

**직업군에 따라 월급의 평균이 다르다.**

5. 백신의 종류와 부작용의 관계를 알아보기 위해 A백신을 접종한 사람 30명, B백신을 접종한 사람 30명을 각각 조사하였다. A백신을 접종한 사람 30명

중에서 18명이 부작용을 경험했고, B백신을 접종한 사람 30명 중 15명이

부작용을 경험했다. 이 데이터를 가지고, 백신의 종류와 부작용 여부가 서로

독립이라고 할 수 있는지 유의수준 5%에서 검정하시오. (5점)

백신과 부작용은 서로 독립이다(A백신과 B백신 간 부작용 비율에 차이가 없다).

백신과 부작용은 서로 독립이 아니다(A백신과 B백신 간 부작용 비율에 차이가 있다).

> # 백신 정보 등록

> *vaccine* = c(rep("A", 30), rep("B", 30))

> # 백신에 따른 부작용 정보 등록

> *side\_effect* = c(rep("Y", 18), rep("N", 12), rep("Y", 15), rep("N", 15))

> # 독립성 검정

> chisq.test(*x*=vaccine, *y*=side\_effect, *correct*=FALSE)

        Pearson’s Chi-squared test

data:  vaccine and side\_effect

X-*squared* = 0.60606, *df* = 1, p-*value* = 0.4363

* p-*value* = 0.4363 > 0.05
* 귀무가설을 기각하지 못한다.
* 백신과 부작용은 서로 독립이다(A백신과 B백신 간 부작용 비율에 차이가 없다).
* A백신과 B백신 간의 부작용 비율의 차이는 통계적으로 유의하다.

6. 교재 p.180 예제 7-3의 데이터를 이용하여, 방과 후 학습 전 성취도

점수와, 방과 후 학습 후 성취도 점수의 산점도를 그리고, 상관계수를

계산하시오. 방과 후 학습 전 성취도 점수를 독립변수로, 방과 후 학습 후

성취도 점수를 종속변수로 하는 회귀직선을 구하시오. (5점

> # 방과 후 학습 전 성취도 점수

> *before* = c(72, 80, 83, 63, 66, 76, 82)

> # 방과 후 학습 후 성취도 점수

> *after*  = c(78, 82, 82, 68, 70, 75, 88)

>

> # 상관계수를 구한다.

> cor(before, after)

[1] 0.921773

상관계수를 계산하시오 < **결과는 9.21773으로 강한 양의 선형관계를 이룬다.** >

>

> *ret* =lm(after~before)

> summary(ret)

Call:

lm(*formula* = after ~ before)

Residuals:

       1        2        3        4        5        6        7

 2.57110 -0.09454 -2.59416  0.06995 -0.42967 -3.76172  4.23905

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  15.4381    11.7438   1.315  0.24572

before        0.8332     0.1567   5.316  0.00315 \*\*

---

Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.022 on 5 degrees of freedom

Multiple R-squared:  0.8497,    Adjusted R-squared:  0.8196

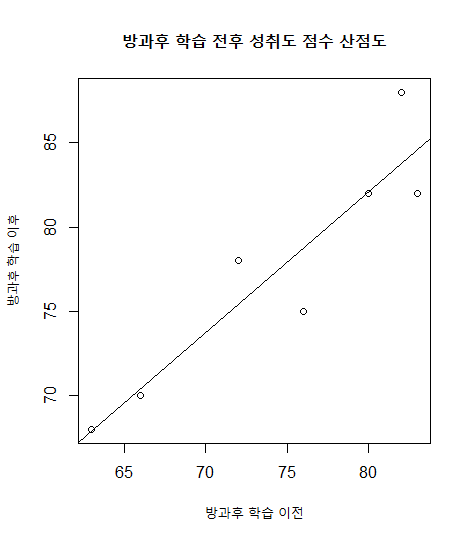
F-statistic: 28.26 on 1 and 5 DF,  p-value: 0.003151

> # 산점도를 그리고,

> plot(before, after, *main*="방과후 학습 전후 성취도 점수 산점도", *xlab*="방과후 학습 이전", *ylab*="방과후 학습 이후")

> # 회귀 직선을 그린다.

> abline(ret)



-- 통계학개론 출석수업 과제 끝 --