

# R프로그래밍및실습 마지막 퀴즈\_답안

2018-06-14

1 아래의 R코드를 이용하여 데이터프레임 health\_2014\_s을 생성하시오.

```
# 홈페이지에서 "건강검진_2014_일부추출.csv" 데이터를 다운로드하여 자신이 지정한 디렉토리에 저장
health_2014 <- read.csv("건강검진_2014_일부추출.csv", fileEncoding = "CP949")
student_id <- 1234567 # <=== 여기에 자신의 학번을 입력할 것 (반드시!!!!)
set.seed(student_id) # 랜덤넘버 생성시 초기값 지정
sidx <- sample(nrow(health_2014), 5000, replace = F) # 인덱스를 랜덤추출
health_2014_s <- health_2014[sidx,] # 추출된 자료
head(health_2014_s)
```

	성별	연령	시도	신장	체중	허리둘레	수축기혈압	이완기혈압	흡연상태	음주여부
5623	1	30	인천	170	65	78	115	70	1	1
7265	2	70	서울	155	55	72	144	87	1	1
9151	1	40	대전	180	80	79	130	72	1	0
297	2	60	서울	165	65	81	100	70	1	1
7673	2	70	대전	160	60	81	109	70	1	1
4799	2	60	대구	170	60	71	123	67	1	1

1.1 위의 R코드를 라인별로 설명하시오.

□ 위의 코드에 코멘트 참조

1.2 health\_2014\_s자료에서 시도별 건강검진 대상자수를 구하라.

```
x <- aggregate(성별~시도, health_2014_s, length)
head(x)
```

시도	성별
1 강원	174
2 경기	1132
3 경남	360
4 경북	283
5 광주	149
6 대구	237

1.3 “성별” 변수를 범주형 변수로, 각 수준(level)값이 1은 “남”, 2는 “여”로 수준이름(label)을 바꾸시오.

```
health_2014_s$성별 <- factor(health_2014_s$성별, labels=c("남", "여"))
head(health_2014_s$성별)
```

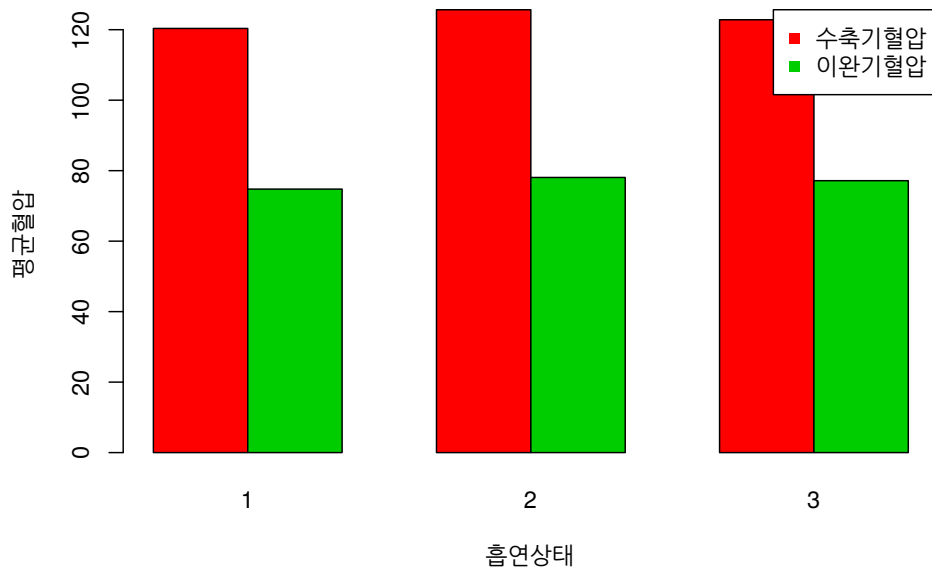
```
[1] 남 여 남 여 여 여
Levels: 남 여
```

```
levels(health_2014_s$성별)
```

```
[1] "남" "여"
```

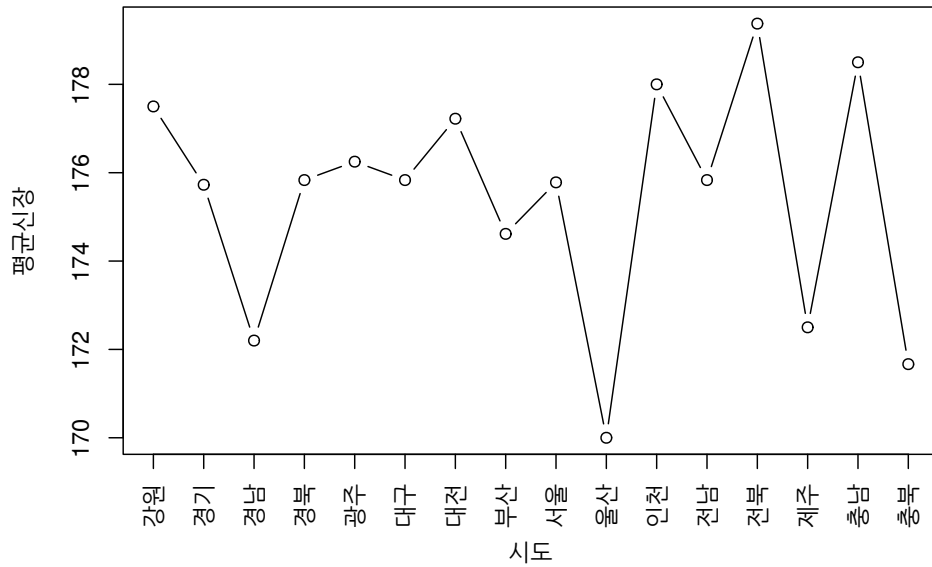
1.4 흡연상태별 수축기혈압과 이완기혈압의 평균을 막대그래프로 나타내시오.

```
m1 <- aggregate(수축기혈압~흡연상태, health_2014_s, mean)
m2 <- aggregate(이완기혈압~흡연상태, health_2014_s, mean)
m <- merge(m1, m2, by="흡연상태")
barplot(t(as.matrix(m[,-1])), beside=T, names.arg = m$흡연상태,
        xlab="흡연상태", ylab="평균혈압", col=2:3)
legend("topright", c("수축기혈압", "이완기혈압"), col=2:3, pch=15)
```



1.5 시도별로 남자 “40대미만”의 평균 “신장”을 “선그래프”로 나타내시오.

```
height <- subset(health_2014_s, subset=(연령 < 40 & 성별=="남"))
m_height <- aggregate(신장~시도, height, mean)
plot(m_height$신장, type="b", xaxt='n', xlab="시도", ylab="평균신장")
axis(side=1, at=seq_along(m_height$시도), labels=m_height$시도, las=2)
```



1.6 “성별”, “연령”에 따라 “신장”의 평균 및 표준편차를 구하시오.

```
m <- aggregate(신장~성별+연령, health_2014_s, mean)
s <- aggregate(신장~성별+연령, health_2014_s, sd)
result <- merge(m, s, by=c("성별", "연령"))
names(result)[3:4] <- c("평균신장", "표준편차")
head(result)
```

	성별	연령	평균신장	표준편차
1	남	30	175.4762	6.375094
2	남	40	176.1618	6.051616
3	남	50	173.6200	5.839505
4	남	60	171.1327	5.931418
5	남	70	168.3862	5.761762
6	남	80	167.3373	5.876596

1.7 흡연상태가 “1”인 여성의 “수축기혈압”과 “이완기혈압”의 상관계수를 구하시오.

```
blood_pressure <- subset(health_2014_s, 흡연상태==1 & 성별=="여")
cor(blood_pressure$수축기혈압, blood_pressure$이완기혈압)
```

[1] 0.7577215

2 아래는 신입생을 대상으로 설문조사한 자료이고 각 변수에 대한 설명은 아래와 같다.

1. 성별: 1.남; 2.여
2. 소속: 단과대학

3. 지원동기: 1. 성적(합격가능성); 2. 장학금; 3. 사회적 평판; 4. 졸업 후 진로; 5. 우수한 교육제도 및 시설; 6. 학문적 흥미와 적성; 7. 학사교류제도
4. 관심사항: 1. 학업; 2. 취업준비; 3. 동아리활동; 4. 교우관계
5. 학교생활만족도: 1. 매우만족; 2. 만족; 3. 보통; 4. 불만족; 5. 매우불만족

```
# 홈페이지에서 "설문조사.csv" 데이터를 다운로드하여 자신이 지정한 디렉토리에 저장
survey_2014 <- read.csv("설문조사.csv", fileEncoding = "CP949")
names(survey_2014)
```

```
[1] "성별"           "소속"           "학교이미지"     "지원동기"       "관심사항"
[6] "학교생활만족도"
```

2.1 소속별/성별 조사대상 학생수를 구하라.

```
survey_2014$성별 <- factor(survey_2014$성별, labels = c("남", "여"))
cnt <- table(소속대학=survey_2014$소속, 성별=survey_2014$성별, useNA = "ifany")
colnames(cnt)[3] <- "무응답"
cnt
```

소속대학	성별		
	남	여	무응답
경영대	227	176	1
과기대	305	132	0
사범대	28	68	0
사회과학대	66	69	1
의대	16	59	2
인문대	180	220	1
자유전공	37	34	0
한의대	48	26	1

2.2 소속별 조사대상 지원동기에 따른 학생수의 비율을 구하라.

```
survey_2014$지원동기 <- factor(survey_2014$지원동기,
                                labels = c("합격가능성", "장학금", "사회적평판", "졸업후진로",
                                             "교육제도및시설", "학문적흥미와적성", "학사교류제도"))
cnt <- table(survey_2014$지원동기, survey_2014$소속, useNA = "ifany")
out <- round(prop.table(cnt, 2), 2)
rownames(out)[8] <- "무응답"
out
```

	경영대	과기대	사범대	사회과학대	의대	인문대	자유전공	한의대
합격가능성	0.34	0.34	0.30	0.33	0.31	0.37	0.44	0.39
장학금	0.03	0.02	0.02	0.04	0.00	0.04	0.00	0.03
사회적평판	0.09	0.06	0.05	0.09	0.03	0.09	0.10	0.01
졸업후진로	0.07	0.13	0.21	0.10	0.23	0.06	0.04	0.12
교육제도및시설	0.01	0.00	0.02	0.01	0.00	0.01	0.00	0.01

학문적 흥미와 적성	0.13	0.14	0.25	0.09	0.31	0.13	0.11	0.37
학사교류제도	0.24	0.25	0.09	0.29	0.01	0.24	0.28	0.00
무응답	0.09	0.06	0.05	0.06	0.10	0.04	0.03	0.07

2.3 소속별 성별 학교생활만족도의 평균점수와 표준편차를 구하라.

```
m_satis <- aggregate(학교생활만족도~소속+성별, survey_2014, mean)
s_satis <- aggregate(학교생활만족도~소속+성별, survey_2014, sd)
satisfaction <- merge(m_satis, s_satis, by=c("소속", "성별"))
names(satisfaction)[3:4] <- c("평균만족도", "표준편차")
satisfaction
```

	소속	성별	평균만족도	표준편차
1	경영대	남	2.351220	0.7026288
2	경영대	여	2.451807	0.6085973
3	과기대	남	2.212766	0.8075137
4	과기대	여	2.222222	0.6915361
5	사범대	남	2.111111	0.6979824
6	사범대	여	2.215385	0.5446417
7	사회과학대	남	2.064516	0.6982641
8	사회과학대	여	2.220588	0.6656965
9	의대	남	1.785714	0.5789342
10	의대	여	2.105263	0.6730583
11	인문대	남	2.201149	0.7751926
12	인문대	여	2.269767	0.6983073
13	자유전공	남	2.111111	0.6666667
14	자유전공	여	2.352941	0.4850713
15	한의대	남	1.521739	0.6232221
16	한의대	여	1.800000	0.6454972