

7장. 자료의 정리1 (일변량 데이터)

7.0.1 자료입력 (직접입력)

```
k_score = c(96, 80, 76, 96, 88, 75, 78, 89, 92, 70) ; k_score
k_score_2 = c(67, 83, 96, 90, 85, 75, 82, 89, 92, 75) ; k_score_2
score=c(k_score, k_score_2) ; score ; mean(score)
```

7.0.2 자료입력 (edit 사용)

```
mydata <- edit(data.frame())
edit(mydata) ; 변수명에 a 기입 ??
a
attach(mydata)
var1
a <- var1 ; mean(a)
```

7.0.3 자료입력(EXCEL 자료) : (강의자료에 Rtest.xlsx)

1. gender(성별) : (1 : 남자, 2 : 여자)
2. eco(주관적 경제력) : (상 : 1, 중 : 2, 하 : 3)
3. iq : (100 - 200)
4. height : (150 - 200)
5. weight : (50 - 100)
6. mid(중간고사성적) : 60 - 100
7. final(기말고사성적) : 60 - 100

[NOTE] 이 파일은 계속 실습용으로 사용할 예정이므로 꼭 보관하시기 바랍니다.

① EXCEL 자료 불러오기

```
EXCEL : d:/Rtest.xlsx 파일 => 다른 이름으로 저장 => Rtest.csv로 저장
data = read.csv("c:/Rtest.csv", header = T) # 변수이름을 인식하여 "data" 에 자료 저장
data
attach(data) # 다시 변수명 할당
weight ; length(weight)
mean = (mid + final) /2 ; mean # 평균 계산
bmi = ( weight / height**2 ) * 10000 ; bmi # 비만도 계산
ls()
```

② 변수 제거

```
ls()
rm(bmi) ; ls() # 변수 (bmi) 제거
rm( list=ls() ) ; ls() # 모든 변수 제거
```

③ 작업공간 저장(R Console 창에서)

```
[파일] => [작업공간 저장] # R Console창에서 : 현재 객체 => Rtest.RDATA로 저장
=> 끝 => 다시 시작 =>
ls() # 확인 되면 아래 생략
attach(data) # 다시 변수명 할당
```

④ Rtest.RDATA => usb에 저장 => 다른 컴퓨터에서 불러오기

```
[파일] => [작업공간 불러오기] # Rtest.RDATA => 불러오기 => 앞으로는 계속 그냥 저장
```

7.1 질적자료(범주형 자료)

7.1.1 분할표(도수분포표)

```
table (gender) # 성별에 대한 도수분포표
table (eco) # 학력에 대한 도수분포표
```

```
gender <- factor(gender, labels = c("남자", "여자") ) # 1 => 남자, 2=> 여자
eco <- factor(eco, labels = c("상", "중", "하" ) )
table (gender) # 성별에 대한 도수분포표
table (eco)
```

7.1.2 막대그래프

```
barplot (table(gender), xlab = "성별", ylab = "도수" ) # 성별에 대한 막대그래프
barplot (table(eco), xlab = "경제력", ylab = "도수" ) # 주관적 경제력에 대한 막대그래프
```

7.1.3 파이그림

```
pie (table(gender), main = "성별") # 성별에 대한 파이그래프
pie (table(eco), main = "경제력") # 주관적 경제력에 대한 파이그래프
```

7.2 양적 자료(숫자형)

7.2.1 줄기-잎-그림

```
stem (height) # 키에 대한 줄기잎그림
```

7.2.2 상자그림

```
boxplot (height, main = "키") # 키에 대한 상자그림
```

7.2.3 히스토그램

```
hist (height, prob = T, main = "우리들의 키") # 키에 대한 히스토그램  
lines ( density(height) )
```

7.2.4 중심경향측도

```
mean (height) # 키에 대한 평균  
median (height) # 키에 대한 중앙값
```

7.2.5 산포의측도

```
var (height) # 키에 대한 분산  
sd (height) # 키에 대한 표준편차  
range (height) # 키에 대한 최댓값, 최솟값 => 범위 = max - min  
IQR (height) # 키에 대한 사분위범위 = Q3 - Q1  
quantile (height) # 키에 대한 사분위수  
summary (height) # 키에 대한 기술통계량
```

7.2.6 표준화 점수

```
scale (height) # 키에 대한 표준화점수  
mean (scale (height))  
var (scale (height))  
summary (scale (height)) # 표준화점수에 대한 기술통계량
```

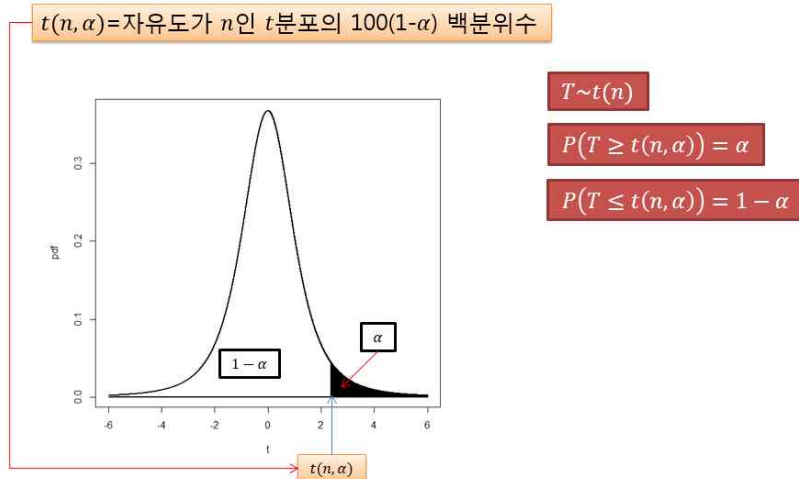
7.2.7 95% 신뢰구간

```
t.test (height) # 키에 대한 95% 신뢰구간
```

7.2.8 95% 신뢰구간 프로그래밍

평균에 대한 95%신뢰구간 $((\bar{x} - t_{\alpha/2}(n-1) * s / \sqrt{n}, \bar{x} + t_{\alpha/2}(n-1) * s / \sqrt{n})$

$t_{\alpha/2}(n-1) = \text{qt}(0.975, n-1)$ # **qt(0.975) : $1-\alpha = 0.975$**



```
n = length(height)
xbar = mean(height)
se = sd(height) / sqrt( n )
lb = xbar - qt(0.975, n-1) * se
ub = xbar + qt(0.975, n-1) * se
cat( " ===== 일표본 신뢰구간 =====", "\n", "\n")
cat( " 표본평균 = ", xbar, " 95% 신뢰구간 ( ", lb, ub, " )" , "\n" )

t.test (height)                # 결과 비교
```

[과제19] (여러분은 일단 따라서 해보시고, 몸무게, bmi 등으로 연습하시길)
연습을 과제로 하시고, 해석을 해주시길..

첨부파일 : 학번이름19.hwp (예 : 20192260홍길동19.hwp)