제2장 확률의 소개

2.1 집합이론

한 실험(조사)의 모든 가능한 결과들의 모임을 그 실험의 **표본공간**(sample space)이라고 한다. 즉, 표본공간은 가능한 결과들의 집합(set)으로 생각할 수 있고, 각각의 결과는 표본공간에 있는 한 원소(element)로 생각할 수 있다. 또한 표본공간의 **부분집합(subset)**을 **사건**(event)이라고 한다.

[예제1] 동전을 **2번** 던지는 실험을 시행할 때, 동전의 앞면을 H, 뒷면을 T로 표시할 때 표본공간 S를 구하여라.

적어도 한 번은 앞면이 나오는 사건 A는 S의 부분집합 $A = \{HH, TH, HT\}$
앞면이 한 번도 나오지 않는 사건 B도 S의 부분집합 $B = \{TT\}$

```
install.packages("prob")    # 'prob' 패키지를 설치
library(prob)    # library()를 이용하여 'prob'를 불러오기
tosscoin(2)   # tosscoin( )을 이용하여 표본공간을 구함
tosscoin(3)
```

subset( ) 함수
조건을 만족하는 벡터, 행렬, 데이터 프레임의 일부를 반환한다.

```
subset(x=, subset=, select= )
subset(x(데이터 이름), # 데이터를 취할 객체(벡터, 행렬, 데이터 프레임)
       subset = (선별 조건) #
       select= c(변수명)  # 데이터 프레임의 경우 선택하고자 하는 열에만 적용됨
        )
```

[예제] 패키지 ggplot2 에 내장 데이터인 **diamonds**를 사용하자.

```
install.packages("ggplot2")
library(ggplot2)
```

str(diamonds)

View(diamonds)

다이어몬드: 신이 흘린 눈물방울[1)]

(1) 조건이 하나 일 때

cut == 'Premium'조건에 만족하는 데이터만 추출함

subset(diamonds, cut=='Premium')

(2) 조건이 두 개 이상 일 때

조건을 두 개 이상 적용할 때, 아래와 같이 나누어 생각할 수 있다.
첫째, 여러 개의 조건 중 하나만 충족해도 된다. (또는)
둘째, 여러 개의 조건을 모두 충족해야 한다. (그리고)

① 또는 (,)

'**또는**' 조건을 적용하기 위해선 **c** 나 **|** (**shift+**)로 묶어주어야 함
묶어주지 않을 경우 쉼표 다음의 조건이 select인 것으로 오인하여 오류가 남

subset(diamonds, c(cut=='Premium', cut=='Fair'))

subset(diamonds, cut=='Premium' | cut=='Fair')

subset(diamonds, cut=='Premium' | color=='E')

② 그리고 (&)

---

1) 4월의 탄생석인 다이아몬드. 다이아몬드의 어원은 '정복할 수 없다'는 뜻의 그리스어 아다마스(adamas)에서 유래되었다. 이미지: pixabay 제공

2) 다이어몬드는 무색에 가까울수록 가치가 높으며, 색상등급은 D(무색)등급부터 Z(옅은 노란색)등급까지 있다.

```
# 여러 조건을 만족시켜야하는 '그리고' 조건은 & 를 사용하여 적용 가능함
subset(diamonds, cut=='Premium' & color=='E')
```

(3) select

(1) 하나의 열만 선택

```
# select로 지정한 열만 반환
subset(diamonds, cut=='Premium' & color=='E', select=clarity)
```

(2) 두 개 이상의 열 선택

```
# c로 묶어주거나 ':'사용
subset( diamonds, cut=='Premium' & color=='E', select=c(clarity, price) )
subset(diamonds, cut=='Premium' & color=='E', select=clarity:price)
```

[예제4] 주사위를 한 번 던져서 나오는 눈의 수를 관찰하는 실험에서 표본공간과 짝수와 홀수 눈이 나오는 경우, 3의 배수가 나오는 경우를 사건으로 표시하여라.

표본공간 S
짝수의 눈이 나오는 경우를 사건 A,
홀수의 눈이 나오는 경우를 사건 B,
3의 배수가 나오는 경우를 사건 C

rolldie(1)

```
( S <- rolldie(1) )  # rolldie( )를 이용하여 표본공간을 구함
  X1
1  1
2  2
3  3
4  4
5  5
6  6
```

```
nrow(S)  # 표본공간 원소의 개수

# subset( ) 함수를 이용하여, X1을 2로 나눈 나머지가 0인 사건 A를 구함
( A <- subset(S, X1 %% 2==0) )  # 참고. X1 %/% 2 X1을 2로 나눈 몫
nrow(A)  # 사건 A의 원소의 개수
nrow(A)/nrow(S)  # 사건 A가 일어날 수학적 확률
```

R에서 문자열(character)은 스칼라, 벡터, 행렬 등으로 저장된다. 문자열을 합쳐서 출력하는 것이 cat( ) 함수[2]이다. 출력할 때 줄바꿈 표시("\n")를 해주자.

```
cat( "P(A)=n(A)/n(S)=", nrow(A)/nrow(S), "\n" )
cat( "사건 A가 일어날 확률은", nrow(A)/nrow(S), "이다.", "\n" )

# X1을 2로 나눈 나머지가 0인 사건 B를 구함
( B <- subset(S, X1 %% 2==1) )

# X1을 3으로 나눈 나머지가 0인 사건 C를 구함
( C <- subset(S, X1 %% 3==0) )
```

[예제5] 1에서 9까지 숫자가 하나씩 적힌 개의 공이 들어 있는 주머니에서 한 개의 공을 꺼내는 시행을 한다. 꺼낸 공에 적힌 숫자가 홀수인 사건을 A, 8의 약수인 사건을 B, 3의 배수인 사건을 C라고 할 때 다음을 구하여라.

(1) $A \cup B$ (2) $A \cap C$ (3) $B \cap C$

(4) A^c (5) B^c (6) $A^c \cap B^c = (A \cup B)^c$

```
# library( ) 함수를 이용하여 'prob'를 불러오기
  library(prob)
# c( )를 이용하여 표본공간을 객체 S에 할당함
  S <- c(1:9)
# A, B, C 사건을 각각 구함
```

2) 사슬 같이 잇는다는 concatenate를 의미하는 함수

```
A=subset(S, S %% 2==1)
B=subset(S, 8 %% S==0)
C=subset(S, S %% 3==0)
```

(1) # union( )을 이용하여 합집합(사건)을 구함

```
union(A, B)
subset(S, S %% 2==1 | 8 %% S==0 )
```

(2) # intersect( )를 이용하여 교집합(사건)을 구함

```
intersect(A, C)
subset(S, S %% 2==1 & S %% 3==0)
```

(3) # intersect( )를 이용하여 교집합을 구하며, integer(0)은 공집합을 나타냄

```
intersect(B, C)
subset(S, 8 %% S==0 & S %% 3==0)
integer(0)
```

(4) # setdiff( )를 이용하여 여집합을 구함

```
setdiff(S, A)
subset(S, S %% 2 != 1)
```

(5) # setdiff( )를 이용하여 여집합을 구함

```
setdiff(S, B)
subset(S, 8 %% S != 0)
```

(6) # setdiff()를 이용하여 여집합을 구함

```
intersect( setdiff(S, A), setdiff(S, B) )
subset(S, S %% 2 != 1 & 8 %% S != 0)
setdiff( S, union(A, B) )  # De Morgan's laws
```