1. sample() 함수\_샘플링

> sample(1:6,10,rep=T,p=c(5,1,10,1,1,1)/10) // sample에 있어야 하는 것

[자료x, size: 샘플링을 구성할 크기, x에서 몇 개를 뽑을지, rep: 뽑은걸 다시 넣는지 안넣는지(복원추출T, 비복원추출F), p=c: 가중치 (어떤 요소가 더 많이 뽑힐지 차이를 둔 것, 숫자 자체가 수학적인 확률은 아님)]

[1] 3 1 4 6 3 3 3 3 3 3

> sample(c("H","T"),10,rep=T)

[1] "T" "T" "T" "T" "T" "H" "H" "T" "H" "H"

2. rep() 함수

> rep(1:4,2) // 1, 2, 3, 4를 2번 반복

[1] 1 2 3 4 1 2 3 4

> rep(1:4, each = 2) // 1, 2, 3, 4를 각각 2번 반복

[1] 1 1 2 2 3 3 4 4

> rep(1:4, c(2,1,2,1)) // 1, 2, 3, 4를 각각 2, 1, 2, 1번 반복

[1] 1 1 2 3 3 4

> rep(1:4, each=2, len=4) // 1, 2, 3, 4를 각각 2번씩 반복하되 길이가 4

[1] 1 1 2 2

> rep(1:4,each=2,len=10)

[1] 1 1 2 2 3 3 4 4 1 1

> rep(1:4, each=2, times=3) // 결과값을 3번 반복

[1] 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4

3. 필터링

> x=c(2,4,1,5,6,7,10)

> y=c(13,2,21,4,6,13,6)

> group=c(2,1,2,1,1,1,2)

> score=c("h","h","l","l","h","l","l")

> ex=data.frame(x,y,group,score)

> ex

x y group score

1 2 13 2 h

2 4 2 1 h

3 1 21 2 l

4 5 4 1 l

5 6 6 1 h

6 7 13 1 l

7 10 6 2 l

> ex1=ex[ex$group==1,] // ex1 = ex에서 group이 1인 자료만 출력

> ex1

x y group score

2 4 2 1 h

4 5 4 1 l

5 6 6 1 h

6 7 13 1 l

4. 필터링\_subset() 함수

> s1=subset(ex,ex$group==1) // 값은 위와 똑같이 나옴

> s1

x y group score

2 4 2 1 h

4 5 4 1 l

5 6 6 1 h

6 7 13 1 l

5. 통계학을 배우는 이유

미지의 세계 (모집단)에서 내포하고 있는 정보를 알기 위해서 표본을 통해 추론한다.

6. 자료의 종류

질적자료: 범주형자료 (categorical data)

: 크기로서의 의미는 없음

: 구분을 위한 기호로서의 의미만 가짐

양적자료: 양을 나타내는 수치자료로 자료의 속성을 그대로 반영함

: 경우에 따라 질적자료의 형태로 변환될 수 있음

Ex 1) 데이터 프레임 ex의 변수 y의 자료 값 4개를 샘플링하여 sy에 저장하라. 단, 변수 y는 데이터 프레임 ex 안의 변수이므로 ex$y라고 해야 한다.

> x=c(2,4,1,5,6,7,10)

> y=c(13,2,21,4,6,13,6)

> group=c(2,1,2,1,1,1,2)

> score=c("h","h","l","l","h","l","l")

> ex=data.frame(x,y,group,score)

> sample(ex$y,4,rep=T,)

[1] 2 13 21 21