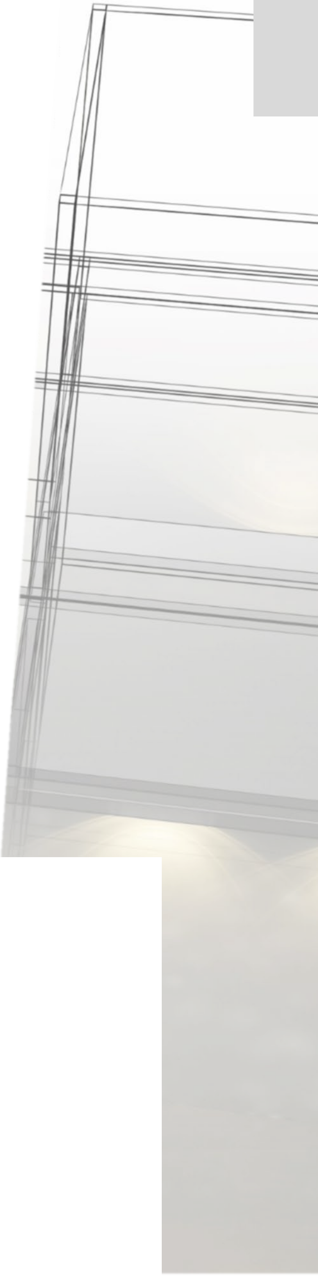
11 강

표본분포 1

통계·데이터과학과
이긍희 교수

학습목표

1. 통계량과 표본분포를 이해할 수 있다.
2. 표본평균의 분포를 이해할 수 있다.
3. 표본비율의 분포를 이해할 수 있다.



01

11강 표본분포 1

통계량과 표본분포

모집단과 표본

- 모집단(population) : 관심의 대상이 되는 전체
- 표본(sample) : 조사 및 측정되는 모집단의 일부
  표본은 모집단을 잘 대표할 수 있도록 임의 추출

모수

◆ 모수(parameter) : 알고 싶은 모집단의 특성값으로 미지의 상수

통계량

◆ 통계량(statistic) : 모수 추정 위한 표본의 함수

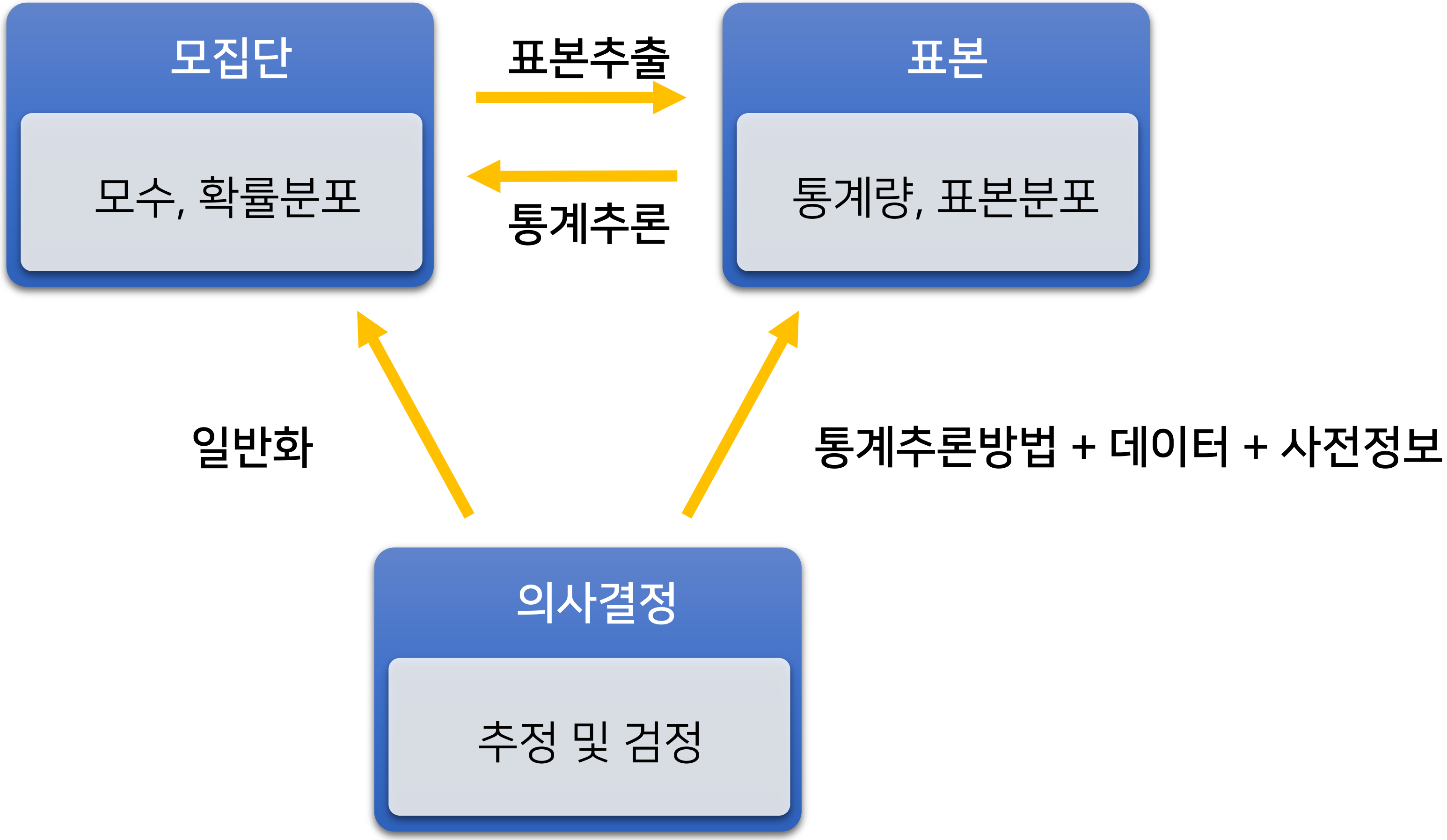
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

표본분포

- 통계량의 분포 : 표본평균, 표본분산 등의 분포

  → 통계적 추정에 이용

통계적 추론


모집단
모수, 확률분포
표본추출
통계추론
표본
통계량, 표본분포
일반화
통계추론방법 + 데이터 + 사전정보
의사결정
추정 및 검정

02

11강 표본분포 1

표본평균의 분포

이산형 확률분포

◆ 주사위의 확률분포표

x	1	2	3	4	5	6	합
$P(X_1 = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

이산형 확률분포

◆ $E(X_1)$

이산형 확률분포

- $Var(X_1)$

표본평균의 분포

◆ $\bar{X} = \frac{1}{2}(X_1 + X_2)$

	1	2	3	4	5	6
1	1	1.5	2	2.5	3	3.5
2	1.5	2	2.5	3	3.5	4
3	2	2.5	3	3.5	4	4.5
4	2.5	3	3.5	4	4.5	5
5	3	3.5	4	4.5	5	5.5
6	3.5	4	4.5	5	5.5	6

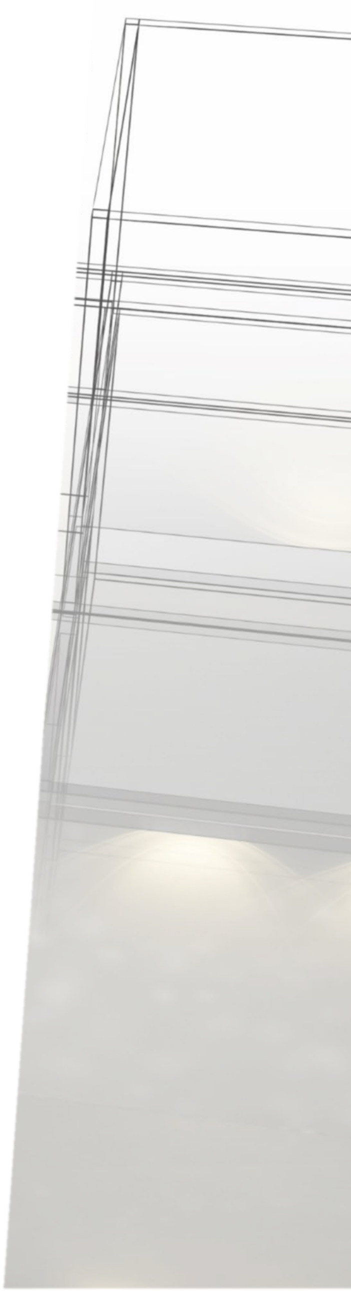
표본평균의 분포

$\bar{X}$	1	1.5	2	2.5	3	3.5
$P(\bar{X})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$

$\bar{X}$	4	4.5	5	5.5	6	합
$P(\bar{X})$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

표본평균의 분포

- $E(\bar{X})$



표본평균의 분포

- $Var(\bar{X})$

표본평균의 분포

◆ 모평균 μ, 분산 σ^2 인 모집단에서 $X_1, X_2, \cdots, X_n$을 임의로 뽑았을 때 표본평균 $\bar{X}$ 의 기댓값과 분산

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

표본평균의 분포

- $E(\bar{X}) = \mu$

표본평균의 분포

- $Var(\bar{X}) = \frac{\sigma^2}{n}$

표본평균의 분포 예

예

평균 10 분산 100인 모집단

(1) 100개의 표본을 임의로 뽑았을 때 표본 평균의 기댓값과 분산은?

표본평균의 분포 예

예

평균 10 분산 100인 모집단

(2) 10,000개의 표본을 임의로 뽑았을 때
표본 평균의 기댓값과 분산은?

표본평균의 분포

◆ $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$의 확률표본일 때

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

표본평균의 분포

◆ $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$의 확률표본일 때

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

표본평균의 분포의 예

예

가구 소득은 평균 300만원, 표준편차 12만원인 정규분포.

(1) 9가구 평균소득의 기댓값은?

표본평균의 분포의 예

예

가구 소득은 평균 300만원, 표준편차 12만원인 정규분포.

(2) 9가구 평균소득의 표준편차는?

표본평균의 분포의 예

예

가구 소득은 평균 300만원, 표준편차 12만원인 정규분포.

(3) 9가구 평균소득이 305만원보다 클 확률은?

03

11강 표본분포 1

표본비율의 분포

지지율

- 1,000명으로 구성된 어느 마을에서 100명을 임의로 뽑아 그 마을의 정책에 대한 지지도를 구함

-100명중 60명이 정책에 찬성

표본비율

◆ 0과 1의 값으로 구성된 모집단에서 n개를 임의로 뽑아서 구한 표본평균

$$\hat{p} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

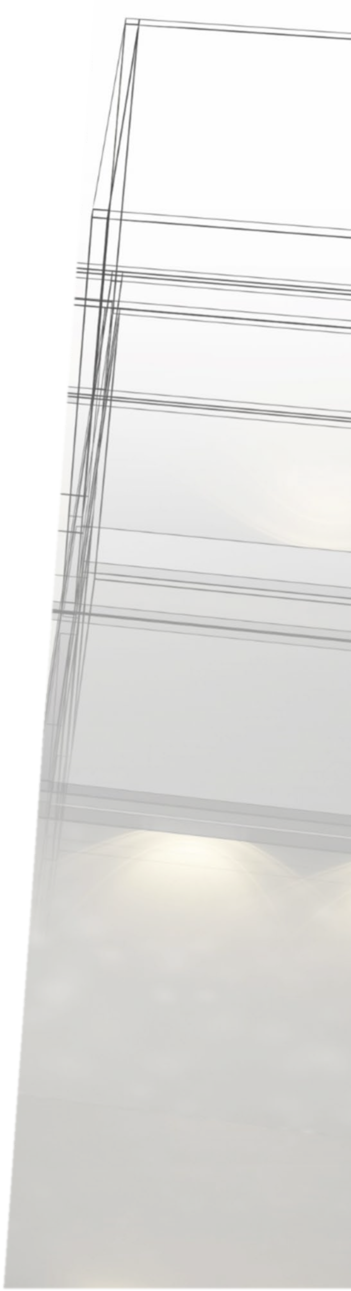
표본비율

- 모집단 A, B, C, D 4명중 A, B가 정책에 찬성하고 C, D가 반대

	A	B	C	D
A	1	1	0.5	0.5
B	1	1	0.5	0.5
C	0.5	0.5	0	0
D	0.5	0.5	0	0

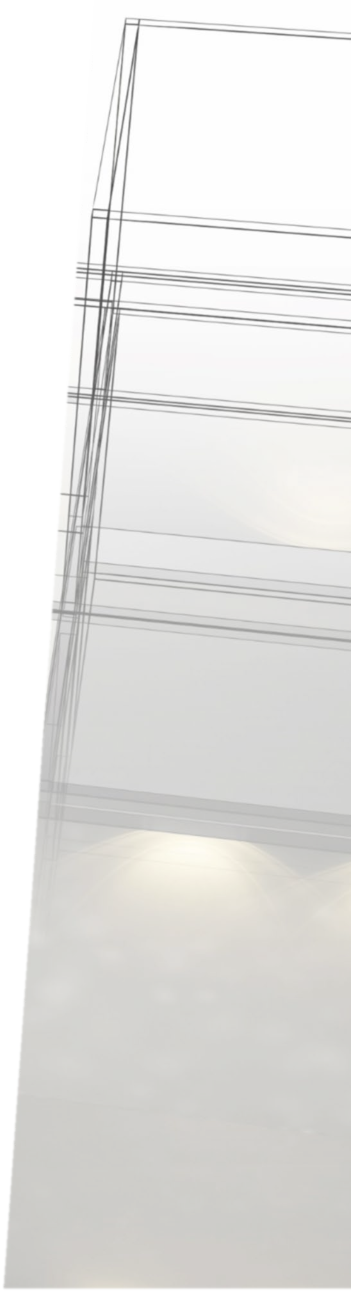
표본비율의 분포

$\hat{p}$	0	0.5	1
$P(\hat{p})$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



표본비율의 기댓값과 분산

- $E(\hat{p})$



표본비율의 기댓값과 분산

- $Var(\hat{p})$

표본비율의 기댓값과 분산

- $E(\hat{p}) = p$

표본비율의 기댓값과 분산

- $Var(\hat{p}) = \frac{p(1-p)}{n}$

표본비율의 수렴

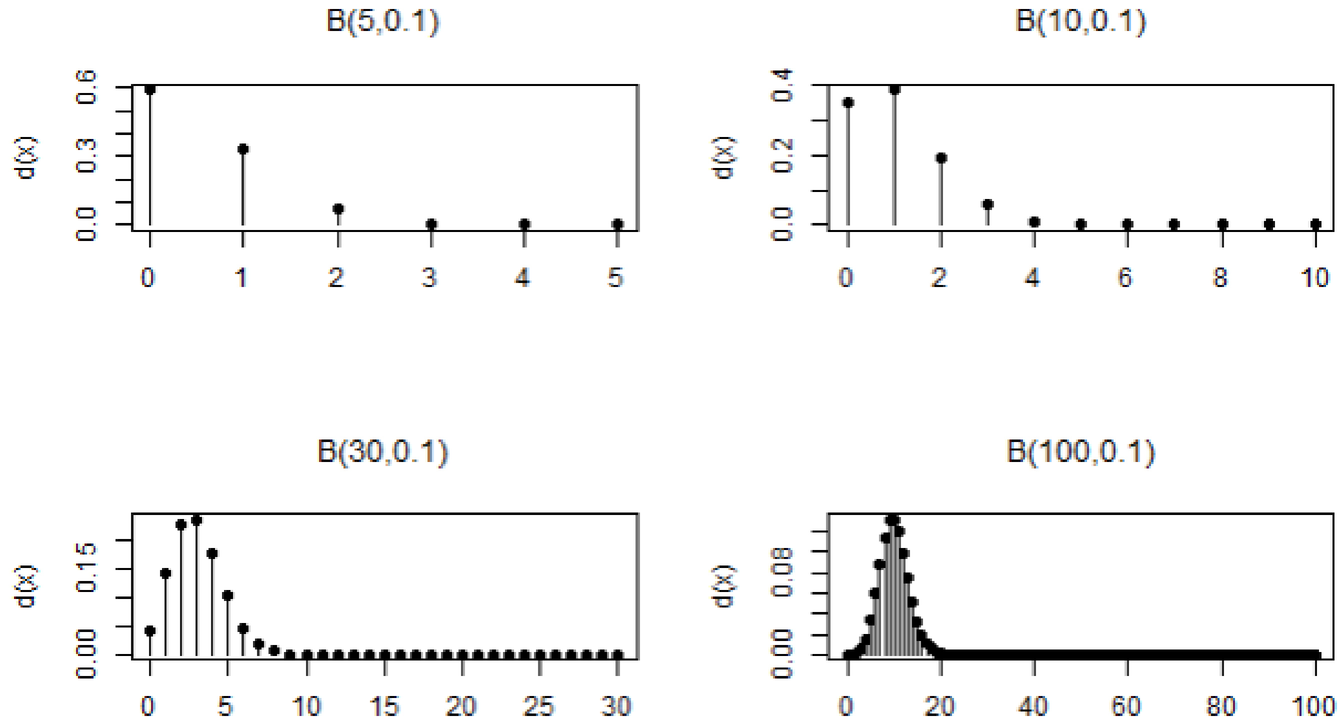
- 대수의 법칙 : 표본수가 커지면 표본비율은 모집단의 비율 p 에 근접

표본비율과 이항분포

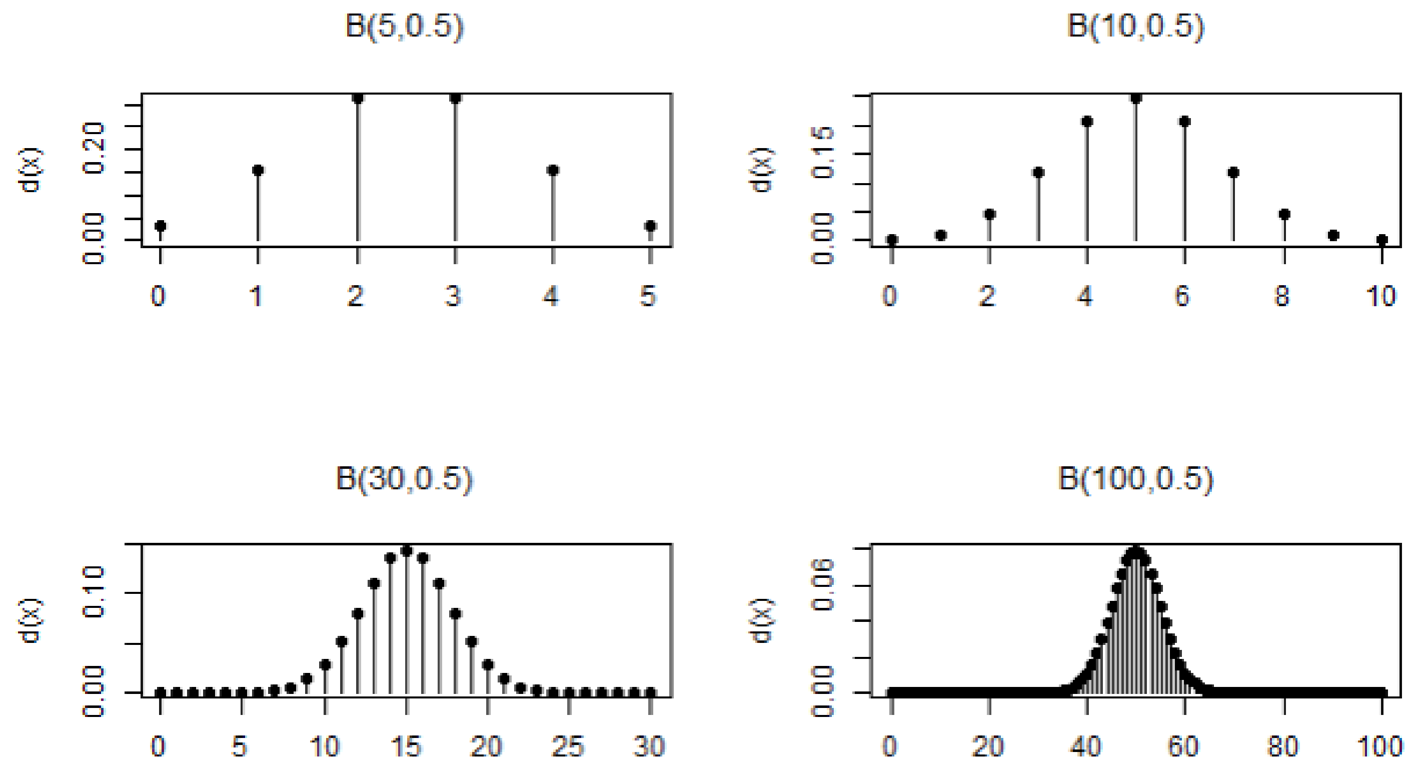
◆ 표본비율 : 이항분포 $B(n,p)$ 따르는 확률변수 $S_n(\sum X_i)$ 을 표본수로 나눈 것

- 표본수 커지면서 S_n 은 정규분포에 근접

이항분포 (p=0.1)


B(5,0.1)
d(x)
0.0
0.3
0.6
0
1
2
3
4
5
B(10,0.1)
d(x)
0.0
0.2
0.4
0
2
4
6
8
10
B(30,0.1)
d(x)
0.00
0.15
0
5
10
15
20
25
30
B(100,0.1)
d(x)
0.00
0.08
0
20
40
60
80
100

이항분포 (p=0.5)



이항분포

- n 이 클 때 $X \sim B(n, p)$는 평균 np, 분산 $np(1-p)$ 인 정규분포에 근접

표본비율의 분포와 연속성 수정

◆ n 이 클 때 표본비율 $\hat{p}$ 은 평균 p , 분산 $\frac{p(1-p)}{n}$ 인 정규분포에 근접 : $np, n(1-p) > 5$

◆ 연속성 수정 :

$$P(a \le X \le b) = P(a - 0.5 \le X \le b + 0.5)$$

이항분포의 예

예

주사위를 180번 던질 때 6이 나오는 횟수가 25회 이상, 35회 이하일 확률은?

이항분포의 예

예

주사위를 180번 던질 때 6이 나오는 횟수가 25회 이상, 35회 이하일 확률은?

표본비율 확률의 예

예

모집단 60세 이상 비율이 0.2인 도시.
100명 추출. 60세 이상 비율이 25%보다
크게 나타날 확률은?

표본비율 확률의 예

예

모집단 60세 이상 비율이 0.2인 도시.
100명 추출. 60세 이상 비율이 25%보다
크게 나타날 확률은?

학습정리

- 표본분포는 통계량의 분포이다.
- 모집단의 확률변수가 정규분포 $N(\mu, \sigma^2)$를 따를 때 표본평균은 정규분포 $N(\mu, \sigma^2/n)$를 따른다.
- 찬성률 p 인 모집단에서 구한 표본의 표본비율 은 근사적으로 정규분포 $N\left(p, \frac{p(1-p)}{n}\right)$를 따른다.

다음시간에는

수고하셨습니다.

11강 표본분포 1

12강 표본분포 2