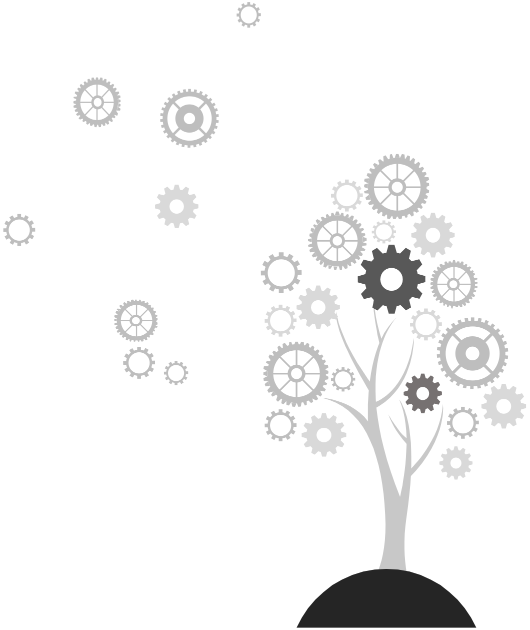
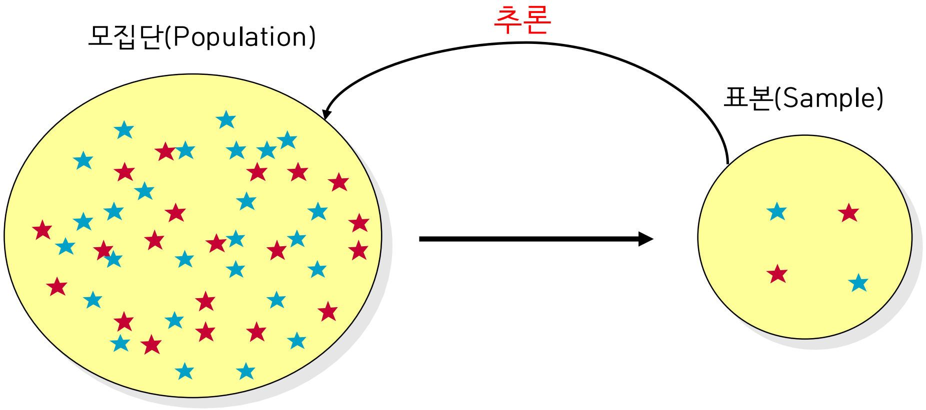
Contents



Chapter 9 표집분포

- 통계학의 목표

: 관측한 자료(모집단의 일부)를 이용하여 모집단에 대한 추측을 하는 것



- 모집단에 대한 모든 자료를 얻는 것(전수조사)은 현실적으로 불가능
- 대신에 모집단의 일부 추출된 자료인 표본을 이용하여 모수에 대해 추론

Chapter 9 표집분포

01 통계량의 확률분포

- 용어

1) 모수 (parameter) : 수치로 표현되는 모집단의 특성 ex. 모평균(μ), 모표준편차(σ)

2) 추론 (inference) : 표본으로 부터 모집단의 성격(모수)을 알아내고자 하는 것

3) 통계량 (statistic) : 표본(관측한 자료)에 의해 결정되는 양 ex. 표본평균($\overline{X}$), 표본분산(S^2)

- 통계량의 특징

1) 통계량의 값은 모수와 통상적으로 같지 않다. ex. $\overline{X} \neq \mu$
2) 통계량은 $X_1, X_2, \ldots, X_n$의 함수이다.
3) 통계량은 확률변수이다.

표집분포

- 통계량의 확률분포 (sampling distribution, 표집분포)

1) 임의표본(random sample) : ① 서로 독립이고 ② 모집단과 같은 분포를 갖도록 모집단으로 부터 임의로 추출된 표본

2) 표집분포 : 통계량의 확률분포

(Ex) 표본평균의 분포

모집단 : $X \sim N(\mu, \sigma^2)$

임의표본1	임의표본2	임의표본3	임의표본4	임의표본5	…
$x_1^{(1)}$	$x_1^{(2)}$	$x_1^{(3)}$	$x_1^{(4)}$	$x_1^{(5)}$	
$x_2^{(1)}$	$x_2^{(2)}$	$x_2^{(3)}$	$x_2^{(4)}$	$x_2^{(5)}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$x_n^{(1)}$	$x_n^{(2)}$	$x_n^{(3)}$	$x_n^{(4)}$	$x_n^{(5)}$	
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
$\bar{x}^{(1)}$	$\bar{x}^{(2)}$	$\bar{x}^{(3)}$	$\bar{x}^{(4)}$	$\bar{x}^{(5)}$	

독립

sample 내 원소들이 (데이터) 표본추출을 할 때마다 바뀌기 때문이다.

표본추출 할 때마다 표본평균(통계량)의 값이 변함

표본평균(통계량)을 확률변수로 간주하고 확률분포를 유도한다

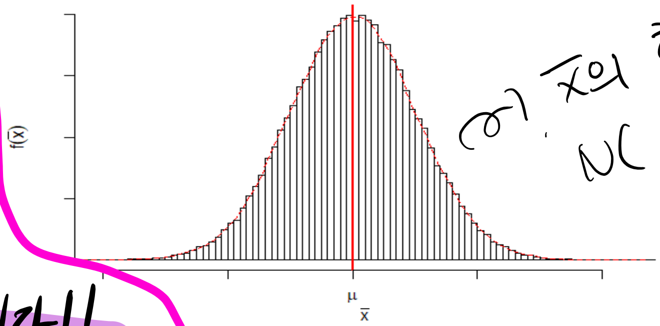
Chapter 9 표집분포

모집단 : $X \sim N(\mu, \sigma^2)$

임의표본 1	임의표본 2	임의표본 3	임의표본 4	임의표본 5	…
$\bar{x}^{(1)}$	$\bar{x}^{(2)}$	$\bar{x}^{(3)}$	$\bar{x}^{(4)}$	$\bar{x}^{(5)}$	

※ $X \sim N(\mu, \sigma^2)$

→ 샘플들이 해당 형태를 띄어야 샘플링이 잘 된 것이다!!



이 $\bar{x}$의 확률분포로 $N(\mu, \frac{\sigma^2}{n})$

Chapter 9 표집분포

- 예제 1

어느 모집단이 세 개의 수 2, 3, 4로 이루어져 있는데 이 수치들은 세 주택의 방의 개수들을 나타내는 것이다. 이 때 X를 각 주택의 방의 개수라고 하면 X는 2, 3, 4 중 하나의 값을 갖게 된다. 세 주택으로부터 두주택을 복원 추출해서 X_1, X_2를 각각 첫번째와 두번째 추출된 주택의 방의 개수라고 할 때, 표본평균 $\bar{X}$의 분포는 X_1, X_2의 분포에 의해 결정되는데 X_1의 분포는 X의 분포와 일치하고, X_2의 분포는 복원추출의 특성상 X_1의 분포와 같이 모집단의 분포인 X의 분포를 따르게 된다. 확률변수 X의 분포는 각 값 2, 3, 4에 1/3씩의 확률을 주는 확률분포를 가지므로 (X_1, X_2)가 취하는 모든 값(x_1, x_2)과 그에 대응하는 $\bar{x}$값과 확률들은 다음과 같이 정리될 수 있다.

(x_1, x_2)	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)	(4,4)
$\bar{x}$	2	2.5	3	2.5	3	3.5	3	3.5	4
확률	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

$\bar{X}$가 취하는 값$(\bar{x})$	2	2.5	3	3.5	4	합계
$f(\bar{x})$	1/9	2/9	3/9	2/9	1/9	1

02 표본평균의 분포와 중심극한정리

- 모평균의 추론

'표본평균'을 사용하여 '모평균'을 추정한다.

1) 추정량 : $\hat{\mu} = \bar{X}$
   cf. 추정량: 모수를 추정하기 위해 만들어진 통계량

$E(X+Y) = E(X) + E(Y)$

2) 표본평균($\bar{X}$)의 기대값과 분산: 모집단의 기대값이 μ, 분산이 σ^2일 때,

① $E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum \mu = \mu$

② $Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

③ $sd(\bar{X}) = \sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$

$\frac{1}{n}E(\Sigma X_i)$
$= \frac{1}{n}\{E(X_1) + \cdots + E(X_n)\}$
$= \frac{1}{n} n\mu = \mu$

3) 정규모집단에서의 표본평균 분포
   : 확률분포가 $X \sim N(\mu, \sigma^2)$인 모집단에서 n개의 표본을 임의로 추출 할 때, 그 표본평균 $\bar{X}$의 분포는 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$이다

$\left\langle E(X) = \mu \rightsquigarrow E(X_i) = \mu,\ i = 1, 2, 3, \cdots, n \right\rangle$

- 예제 2

Chapter 9 표집분포

만약, 모집단이 정규분포를 따른다는 정보가 주어지어 있으면, 중심극한 정리를 사용하지 않는다.

- 중심극한정리(central limit theorem)

: 모집단의 평균이 μ이고 분산이 σ^2일 때, 임의로 추출된 표본의 표본평균 $\bar{X}$ 는 표본의 크기 n이 큰 경우 $(n > 30)$ 근사적으로 정규분포를 따르게 되며 그 평균은 μ이고 분산은 $\frac{\sigma^2}{n}$가 된다.

← 한 번의 샘플링을 통한 표본

모평균!

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

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

해당 표시가 중요하다.

n이 크면, 표본평균의 평균과 모평균이 같아진다 $(E(\bar{X}) = \mu)$

이를 표준화 하면 다음과 같다

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

- 예제 3

① ②

$X \quad P(X = x) = f(x)$

얘를 ②로 바꿔쓸 수 있다.