

## 6. 확률 표본과 추정

## 6.1 확률 표본과 중심 극한 정리

### 확률 표본 (Random Sample)

모집단을 구성하는 모든 구성원들이 표본으로 뽑힐 확률이 동일하게 되도록 추출한 표본

Random이란 무작위라는 의미, 모든 구성원이 뽑힐 확률이 동일하다는 것과 동일한 개념

## 6.1 확률 표본과 중심 극한 정리

### 중심 극한 정리 (Central Limit Theorem)

모집단의 분포를 모르는 경우, 모수의 추정량에 대한 확률계산이 불가능했으나 모집단의 분포와 상관없이 **표본 평균을 이용하면**

**정규 분포를 활용**하여 확률의 계산이 가능

## 6.2 점 추정

### 추정 (Estimation)

추정이라는 것은 모르는 것, 미지의 숫자, 모수 등이 이것일 것이라고 얘기해주는 것

통계학에서는 모르는 것을 추정하는 주제로는 점 추정, 구간 추정, 분포 추정으로 구분

## 6.2 점 추정

### 점 추정 (Point Estimation)

관심의 대상은 모수이며, 모수를 추정하는 통계량이 추정량이므로, **모수가 어떤 특정 값일 것이다**라고 추정

## 6.2 점 추정

### 점 추정을 하는 추정량이 가져야 할 성질

불편성 (Unbiasedness) :  $E(\hat{\theta}) = \theta$

최소 분산성 (Minimum Variance) :  $\hat{\theta}$ 가 여러 개 있을 때 분산이 제일 작은 것을 선택

cf. 일반적으로 표본 평균이 대표값인 이유는 표본 평균은 불편 추정량을 만족하지만 중위수나 최빈값은 동일하지 않기 때문

## 6.3 구간 추정

### 구간 추정 (Interval Estimation)

모평균과 표본 평균이 일치할 확률은 매우 낮을 수 있으며 이를 보완하기 위해

특정 값이 아닌 일정 구간에 속할 확률로 표현할 필요가 있음

구간추정은 신뢰구간 (Confidence Interval)과 같은 의미

## 6.3 구간 추정

모수  $\theta$  의 95% 신뢰구간이란?

크기가 동일한 100개의 서로 다른 표본에서 동일한 구간 계산 공식에 따라 100개의 신뢰구간을 구하면

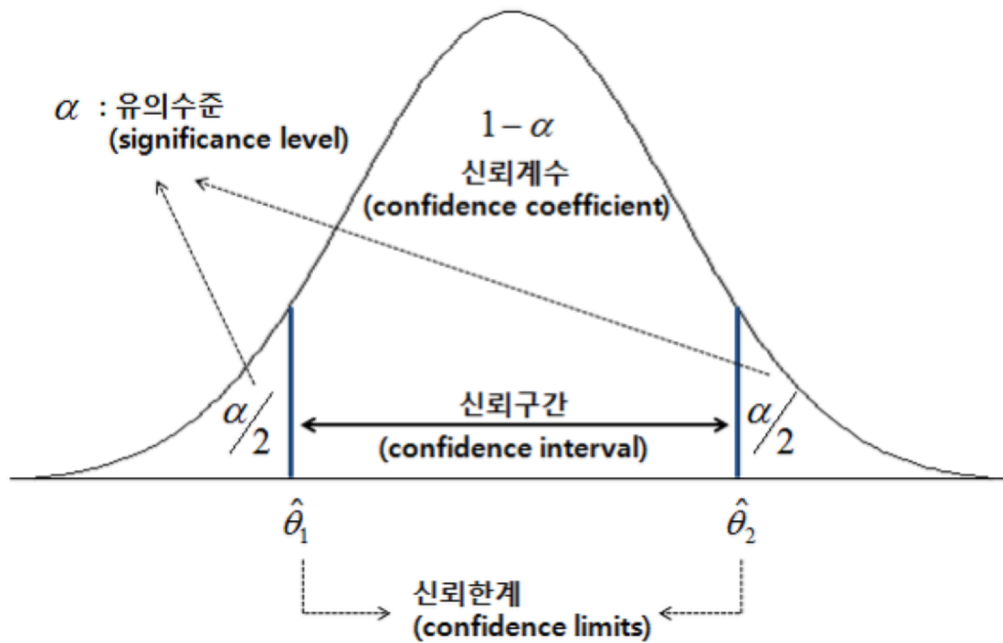
이 가운데 95개의 신뢰구간이 모수  $\theta$ 를 포함한다는 것을 의미

$$P(a < \theta < b) = 0.95 = 95\%$$



## 6.3 구간 추정

### 신뢰구간 변환



$$1 - \alpha = P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$