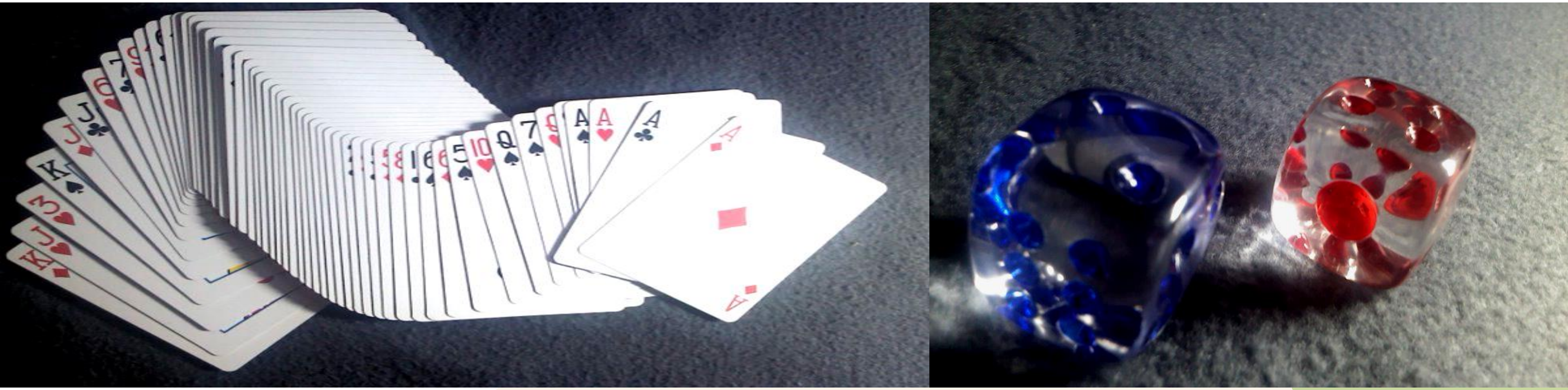
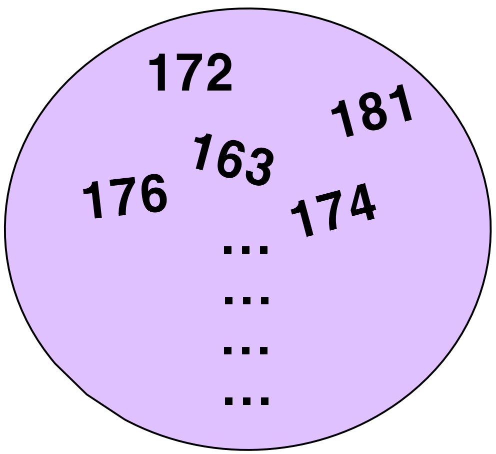
# Chapter 10. 모집단과 표본이론



# 10. 1 개　요

❖ 모집단 : 관심 있는 연구 대상의 특성에 대한 모든 관측치의 집합

❖ Ex)




충남대 남학생 키 전체의 집합

평균은 $\mu$ , 분산은 $\sigma^2$ => 모수

분포 형태는 정규분포 => 가정

임의로 고른 충남대 남학생의 키를 $X$ 라 하면 $X \sim N(\mu\ , \sigma^2)$ 이다

# 10. 2 모집단과 표현

- 모집단을 분포로 표현 할 수도 있다.
  이 경우, 이러한 분포를 따르는 확률 변수 $X$ 가
  존재하게 되며, 관찰되는 관찰치를 그 순서에 따라

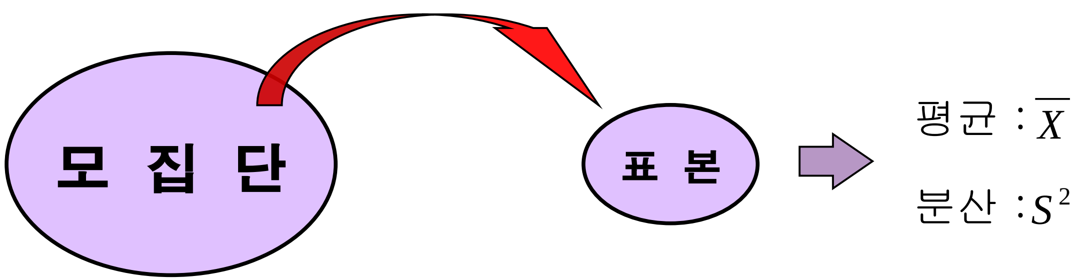
$$X_1, X_2, \cdots, X_n$$

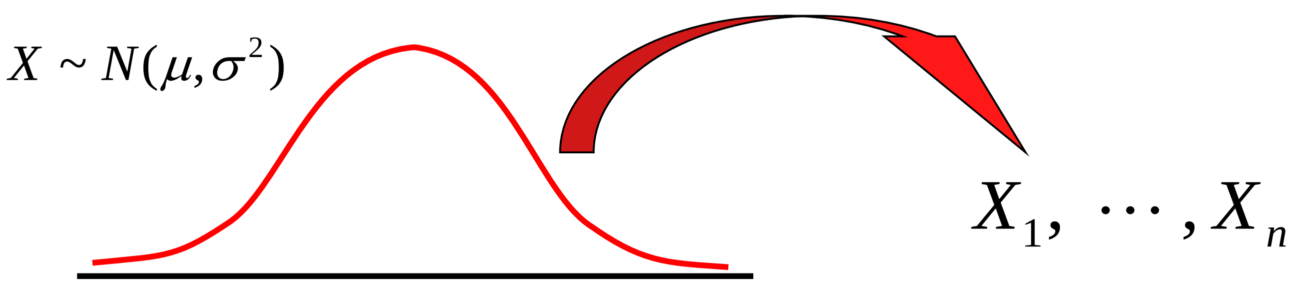
으로 표현하게 된다

  - 모집단의 부분집합을 표본이라 한다
    특히 모집단의 모든 원소들이 표본으로 뽑힐 가능성이
    같도록 보장된 표본을 확률표본( RandomSample)
    이라 한다

# 10. 3 확률표본과 표본분포

❖ 모집단을 확률변수 $X$의 분포라 하면 서로 독립이며 모집단 분포를 따르는 $X_1, X_2, \cdots, X_n$들을 크기가 $n$인 확률표본이라 한다



# 통계량(statistic)

❖ 표본으로부터 계산되는 표본의 특성치를 통계량(statistic)이라 한다. 이는 $X_1, \cdots, X_n$의 함수이다

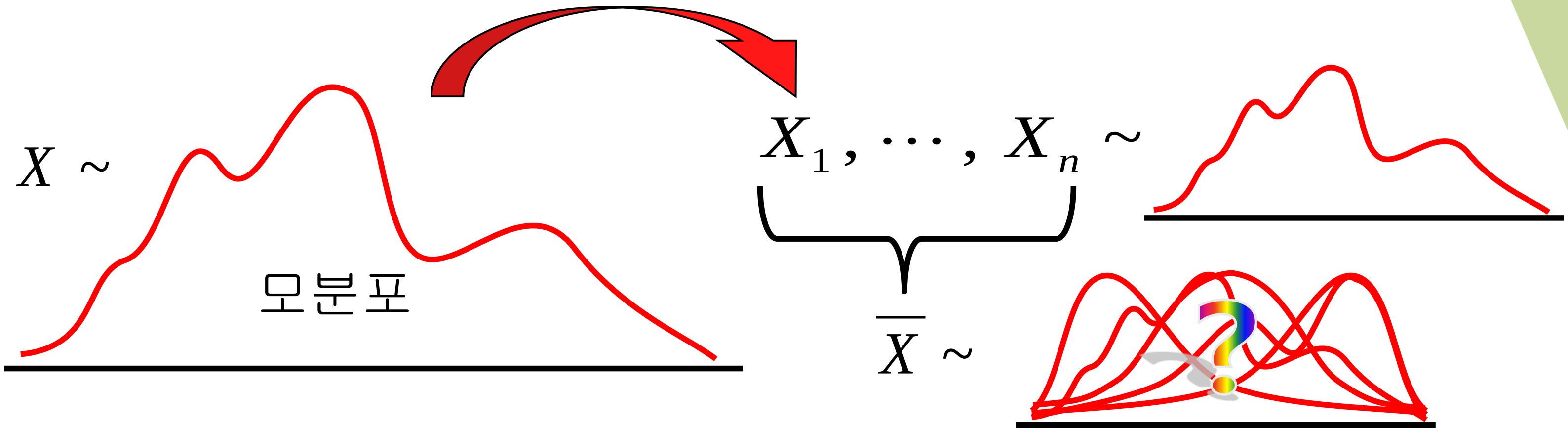
- Ex) $\overline{X}$ , $S^2$ , $\max$

표본분산

표본평균

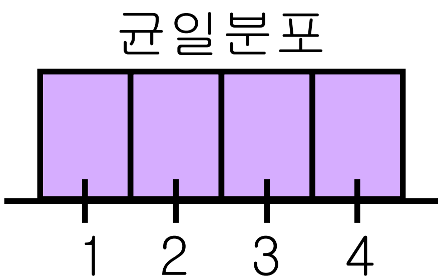
❖ 통계량들이 갖는 확률분포를 표본분포(sampling distribution)라 한다

# 10. 4 표본평균의 분포




❖ Ex) 유한 모집단 $X = \{1, 2, 3, 4\}$

$$X \sim f(x) = \frac{1}{4} \ , \ x = 1, 2, 3, 4$$




크기가 2인 복원 확률표본을 뽑는다.

이때 $\bar{X}$의 분포는?

# 크기가 2인 복원 확률표본평균의 분포

| 가능한 표본 | prob | $\overline{X}$ |
|---|---|---|
| 1 , 1 | $1/16$ | 1 |
| 1 , 2 | $1/16$ | 1.5 |
| 1 , 3 | $1/16$ | 2 |
| 1 , 4 | $1/16$ | 2.5 |
| 2 , 1 | $1/16$ | 1.5 |
| 2 , 2 | $1/16$ | 2 |
| 2 , 3 | $1/16$ | 2.5 |
| 2 , 4 | $1/16$ | 3 |
| 3 , 1 | $1/16$ | 2 |
| 3 , 2 | $1/16$ | 2.5 |
| 3 , 3 | $1/16$ | 3 |
| 3 , 4 | $1/16$ | 3.5 |
| 4 , 1 | $1/16$ | 2.5 |
| 4 , 2 | $1/16$ | 3 |
| 4 , 3 | $1/16$ | 3.5 |
| 4 , 4 | $1/16$ | 4 |

## 모평균

$$\mu = \frac{1+2+3+4}{4} = 2.5$$

$$or \quad 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4} + 4\cdot\frac{1}{4} = 2.5$$

## 모분산

$$\sigma^2 = E(X^2) - E(X)^2$$

$$= (1^2+2^2+3^2+4^2)\cdot\frac{1}{4} - \left(\frac{10}{4}\right)^2$$

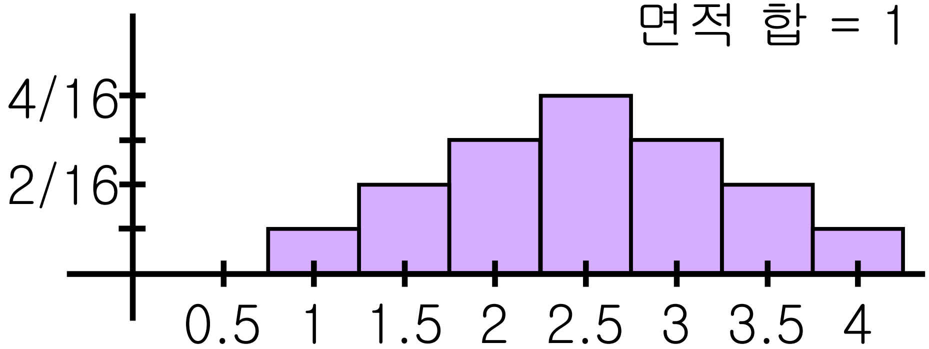
$$= \frac{30}{4} - \frac{25}{4} = \frac{5}{4}$$

$$\therefore \sigma = \sqrt{\frac{5}{4}} = 1.118$$

# 크기가 2인 복원 확률표본평균의 분포

$\overline{X}$ 의 분포

| $\bar{x}$ | $p(\bar{x})$ |
|---|---|
| 1 | $1/16$ |
| 1.5 | $2/16$ |
| 2 | $3/16$ |
| 2.5 | $4/16$ |
| 3 | $3/16$ |
| 3.5 | $2/16$ |
| 4 | $1/16$ |



# 크기가 2인 복원 확률표본평균의 평균과 분산

❖ $\overline{X}$의 기대값 (평균)

$$E(\overline{X}) = \mu_{\overline{X}} = 1\cdot\frac{1}{16} + 1.5\cdot\frac{2}{16} + \cdots + 4\cdot\frac{1}{16}$$
$$= 2.5 = \mu$$

❖ $\overline{X}$의 분산

$$\sigma_{\bar{x}}^{2} = E(\overline{X}^{2}) - [E(\overline{X})]^{2}$$
$$= 1^{2}\cdot\frac{1}{16} + 1.5^{2}\cdot\frac{2}{16} + \cdots + 4^{2}\cdot\frac{1}{16} - (2.5)^{2}$$
$$= 0.625 = \frac{5}{8} = \frac{5}{4}\cdot\frac{1}{2} = \frac{\sigma^{2}}{n}$$

$$\sigma^{2} = \frac{5}{4}$$
$$n = 2$$

# 표본평균의 평균과 분산

- 평균 $\mu$ , 분산 $\sigma^2$ 을 갖는 모집단으로부터 크기가 $n$ 인 표본을 복원추출하면

$$\mu_{\overline{X}} = \mu \ , \quad \sigma_{\overline{X}}^{\ 2} = \frac{\sigma^2}{n}$$

- $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ 은 표본평균의 표준편차로 특별히 표준오차(Standard error)라 불린다

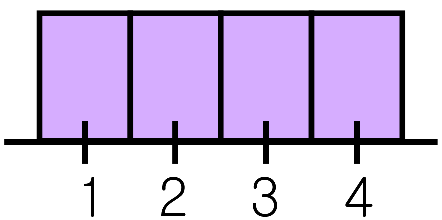
- $\overline{X}$ 는 $\mu$ 의 추정에 사용되며 추정에서 발생하는 편차(Deviation)를 오차(Error)라 부른다

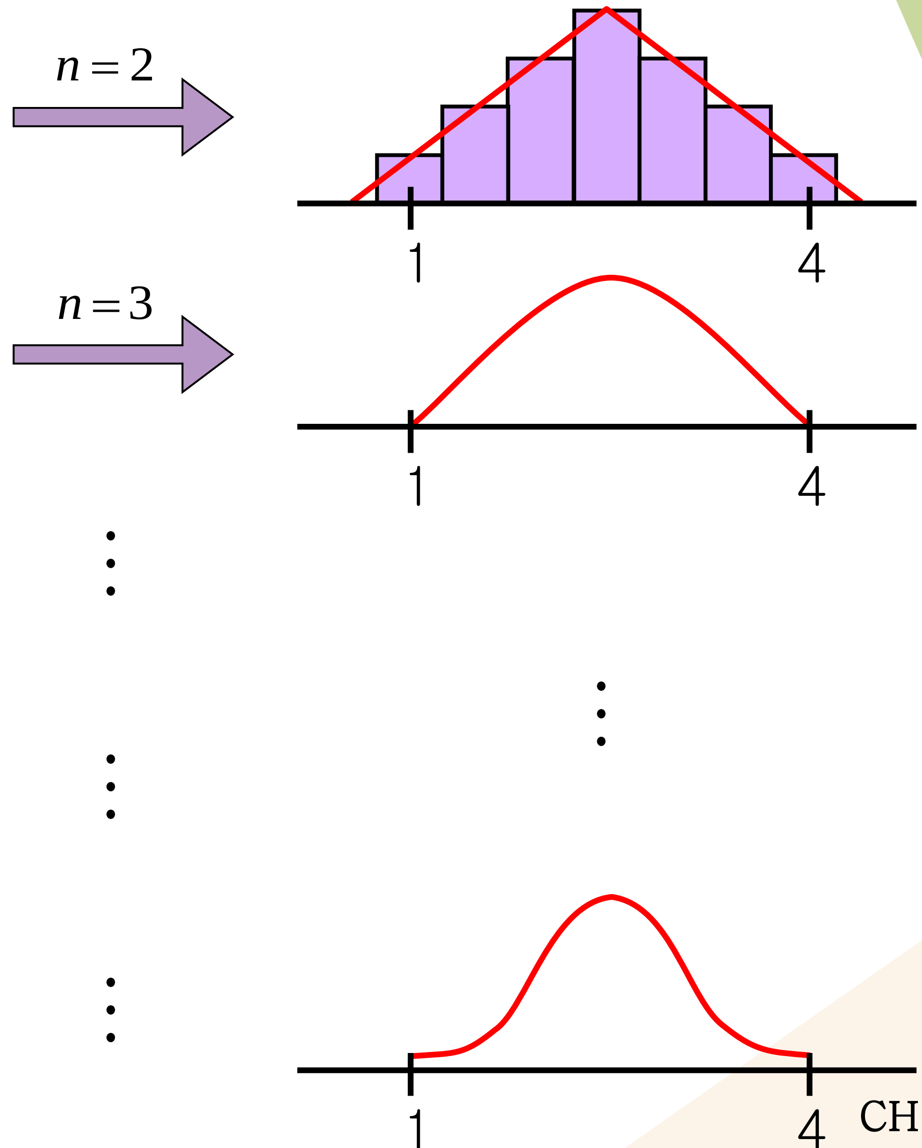
# 표본평균의 분포형태

$\overline{X}$의 분포형태는?

$$X \sim f(x) = \frac{1}{4}$$

$$x = 1, 2, 3, 4$$




모집단이 원래 정규분포면

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

# 중심극한정리

- 중심극한정리(central limit theorem)
  - 평균이 $\mu$이고 분산이 $\sigma^2$인 임의의 모집단으로부터 뽑은 확률표본의 평균 $\overline{X}$의 분포는 표본의 크기가 크면 근사적으로 평균 $\mu$, 분산 $\frac{\sigma^2}{n}$인 정규분포 $N(\mu, \frac{\sigma^2}{n})$에 따른다.

    즉,

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

$NOTE$. 중심극한정리는 $\overline{X}$의 분포형태에 관한 정리임

# 확률예제

- Ex) 고교생 1인당 월평균 사교육비는 500,000원, 표준편차는 80,000원이라 한다. 64명의 고교생을 임의로 뽑았을 때 월 사교육비의 평균이 482,000원을 초과할 확률은?

$$\underbrace{X_1, \cdots, X_{64}} \sim ?\ (\mu = 500{,}000,\ \sigma = 80{,}000)$$

분포의 형태는 모름

$$\overline{X} \sim N(\mu_{\overline{X}} = 500{,}000,\ \sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{80{,}000}{\sqrt{64}} = 10{,}000)$$

즉, $\overline{X} \sim N(500{,}000\ ,\ 10{,}000^2)$

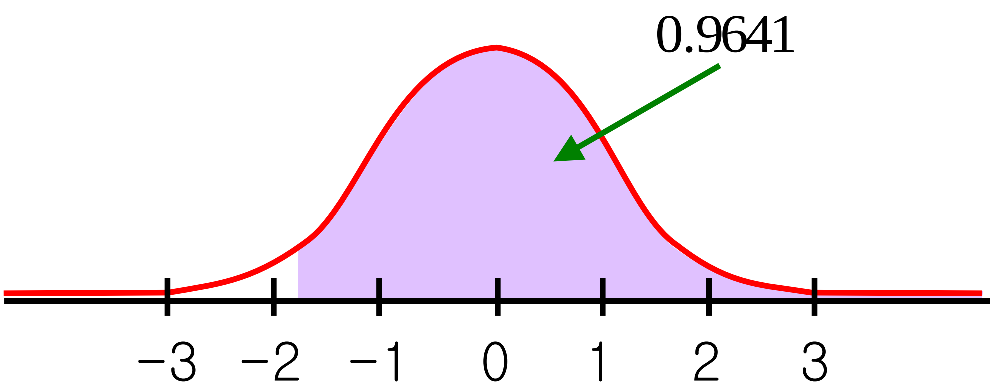
# 확률예제

$$P(\overline{X} > 482{,}000) = P\left( \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{482{,}000-500{,}000}{10{,}000} \right)$$

$$= P(Z > -1.8)$$

$$= 0.5 + P(0 < Z < 1.8)$$

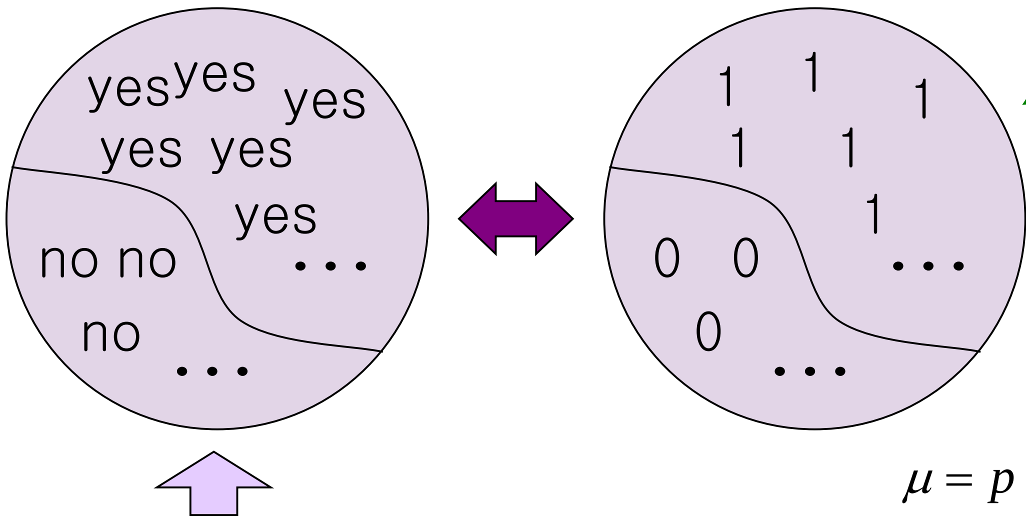
$$= 0.5 + 0.4641 = 0.9641$$



# 10. 5 표본비율의 분포

❖ 충남대남학생의 공주대와의 통합에 대한 의견의 집합(질적 모집단)

Yes를 1, no를 0으로 표현




모집단 크기 $N$

모비율 $p$ = 모집단 내 yes의 비율

$$= \frac{1+1+\cdots+1}{N} = \text{Yes를 1, no를 0으로 본 평균}$$

$$\mu = p$$

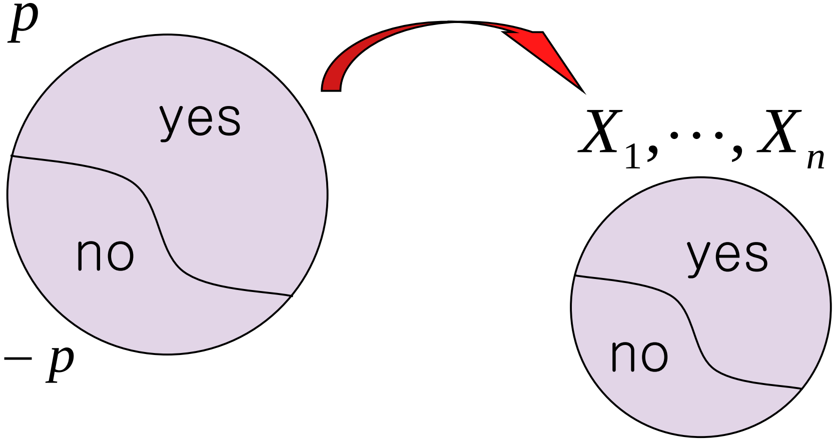
$$\sigma^2 = E(X^2) - [E(X)]^2$$

$$= \frac{1^2+1^2+\cdots+1^2}{N} - p^2 = p(1-p)$$

# 표본비율

$NOTE.$ 모비율도 모평균의 일종

$p$




$q = 1 - p$

$X_1, \cdots, X_n$

표본내 yes의 수 $= X$
$(= \sum X_i$ , yes를 1
n0를 0)

$$X \sim b(n, p)$$

$$\frac{X}{n}(\text{표본비율}) = \frac{\sum X_i}{n} = \overline{X} \;\dot\sim\; N(\mu_{\overline{X}} = \mu = p \,,\, \sigma_{\overline{X}}{}^2 = \frac{\sigma^2}{n} = \frac{pq}{n})$$

# 표본비율의 정규근사

$$\frac{X}{n}(\text{표본비율}) = \frac{\sum X_i}{n} = \overline{X} \stackrel{\cdot}{\sim} N(\mu_{\overline{X}} = \mu = p\ ,\ \sigma_{\overline{X}}{}^2 = \frac{\sigma^2}{n} = \frac{pq}{n})$$

즉, $$\frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} \stackrel{\cdot}{\sim} N(0,1)$$

또는, $$\frac{X - np}{\sqrt{npq}} \stackrel{\cdot}{\sim} N(0,1)$$

또는, $X \stackrel{\cdot}{\sim} N(np\ ,\ npq)$

※ $X \sim b(n, p)$이며 $n$이 커지면 $X \stackrel{\cdot}{\sim} N(np, npq)$

# 확률예제

❖ Ex)모비율 $p=0.4$인 모집단에서 크기가 $n=625$인 표본을 뽑았을때 표본비율이 0.38에서 0.42 사이에 있을 확률은?

$$P\left(0.38 < \frac{X}{n} < 0.42\right) \qquad \frac{X}{n} \sim N(p, \frac{pq}{n})$$

$$\approx P\left(\frac{0.38-0.4}{\sqrt{0.4\times0.6/625}} < Z < \frac{0.42-0.4}{\sqrt{0.4\times0.6/625}}\right)$$

$$= P\left(\frac{-0.02}{0.0196} < Z < \frac{0.02}{0.0196}\right)$$

$$= P(-1 < Z < 1) = 2\cdot P(0 < Z < 1) = 2\times0.3413 = 0.6826$$