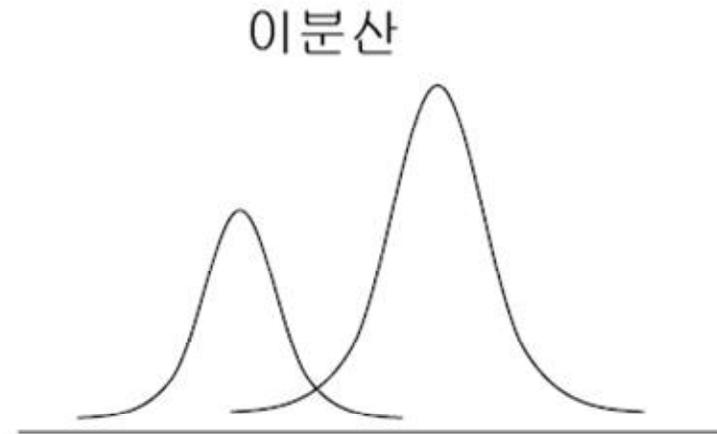
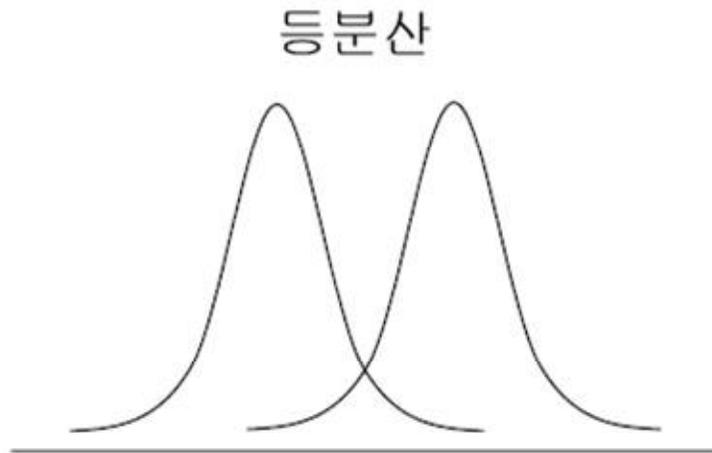


## 1. 등분산, 이분산

- 분산(Variance) : data의 퍼져있는 정도
- 등분산 (Homogeneity of variance) : 2개의 모집단(Population) 에서 추출된 각 sample 간의 분산이 **같음**
- 이분산 (Heteroscedasticity of variance) : 2개의 모집단(Population) 에서 추출된 각 sample 간의 분산이 **다름**



## t-분포 ( t - distribution)

- ① 정의 : 임의의 모집단  $X \sim N(\mu, \sigma^2)$  에서 추출한 크기가  $n$ 인 랜덤포본(sample)이  $X_1, \dots, X_n$  일 때

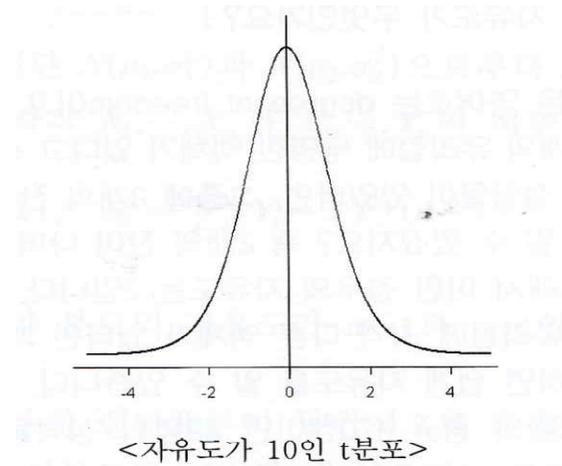
표준 평균 :  $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$

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

➔ 자유도 (Degree of freedom)가  $n-1$ 인 t-분포  $T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$

- ② 표기 :  $X \sim t_{(n-1)}$

- ③ 분포의 형태 :



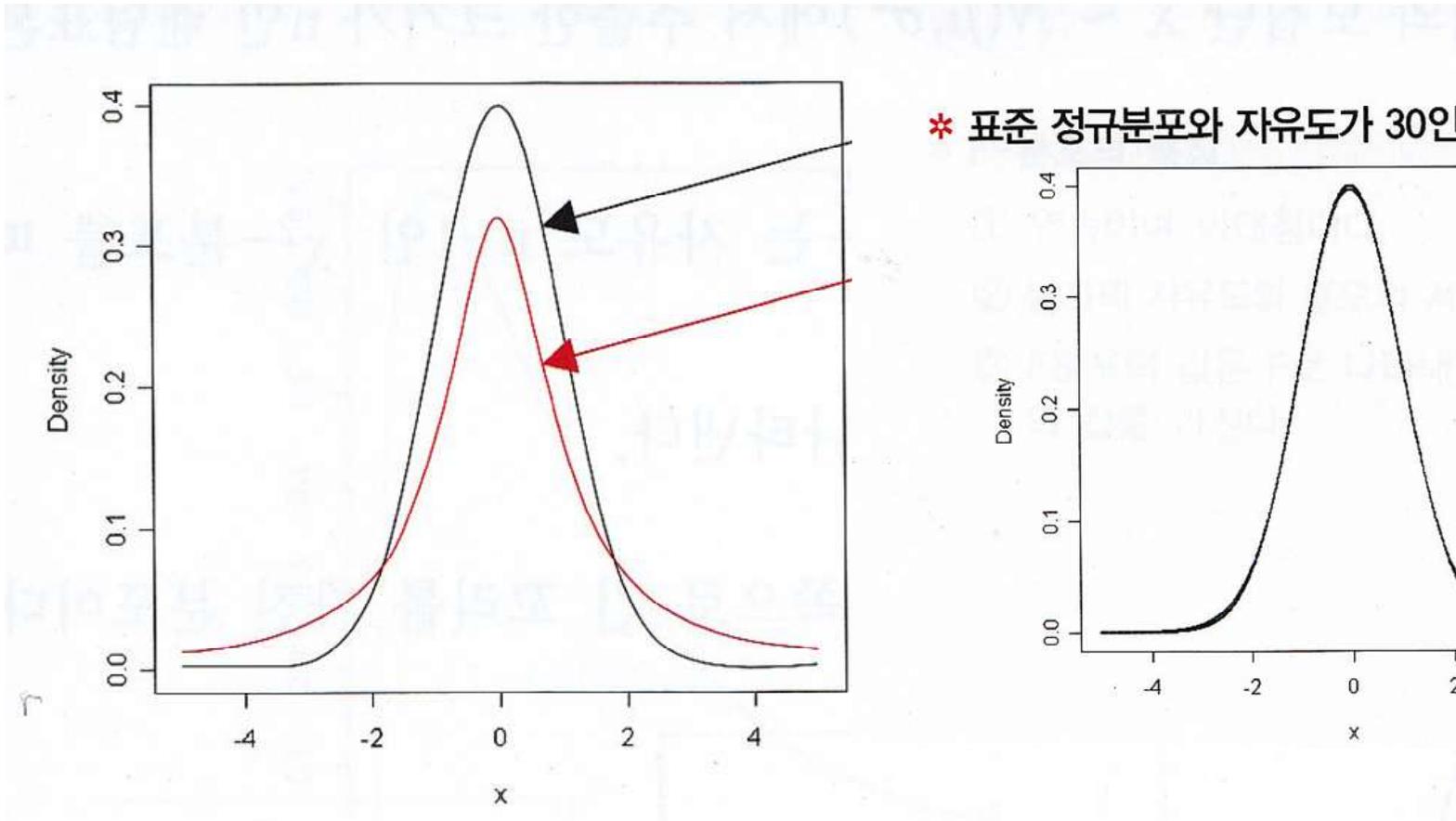
- ④ 평균 & 분산 :

$$E(X) = \mu = 0$$

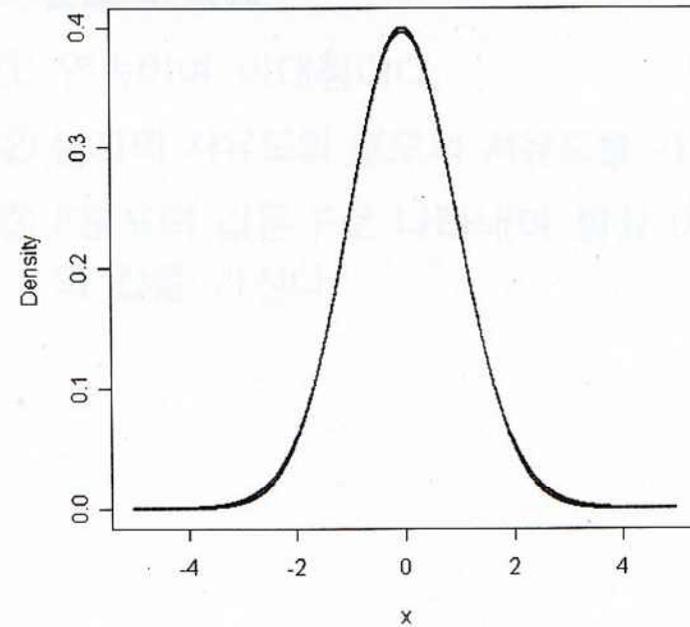
$$Var(X) = df / (df - 2)$$

# t-분포 ( t - distribution)

< 표준정규분포와 t-분포의 비교 >



\* 표준 정규분포와 자유도가 30인 t-분포 그래프



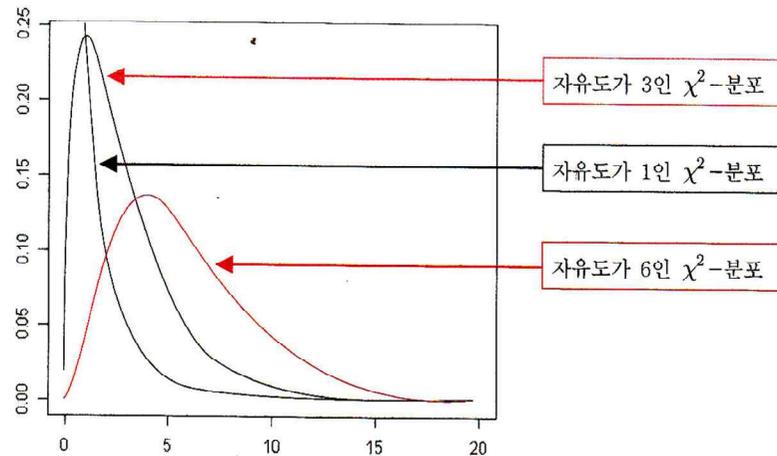
# 카이-제곱 분포 (Chi-square distribution)

- ① 정의 : 정규분포를 따르는 변수의 분산에 대한 신뢰구간을 구할 때 사용  
임의의 모집단  $X \sim N(\mu, \sigma^2)$  에서 추출한 크기가  $n$  인 랜덤표본이  $X_1, \dots, X_n$  일 때

$$\Rightarrow \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1) S^2}{\sigma^2} \text{ 는 d.f. } n-1 \text{ 인 } \chi^2 - \text{분포를 따름}$$

- ② 표기 :  $\frac{(n-1) S^2}{\sigma^2} \sim \chi^2(n-1)$

- ③ 분포의 형태 :



## F-분포 ( F - distribution)

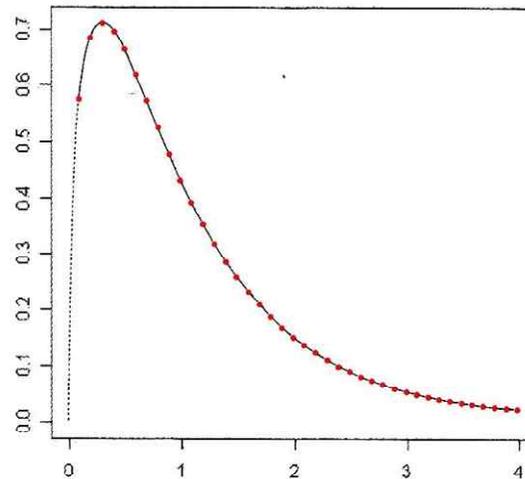
- ① 정의 : 서로 독립인 두 정규모집단  $N(\mu_1, \sigma_1^2)$  과  $N(\mu_2, \sigma_2^2)$  으로부터 크기가  $m$  과  $n$  인 랜덤표본을 추출하여 각각  $X_1, \dots, X_m$  과  $Y_1, \dots, Y_n$  라 하면

$$S_1^2 = \sum_{i=1}^m (X_i - \bar{X})^2 / (m-1), \quad S_2^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)$$

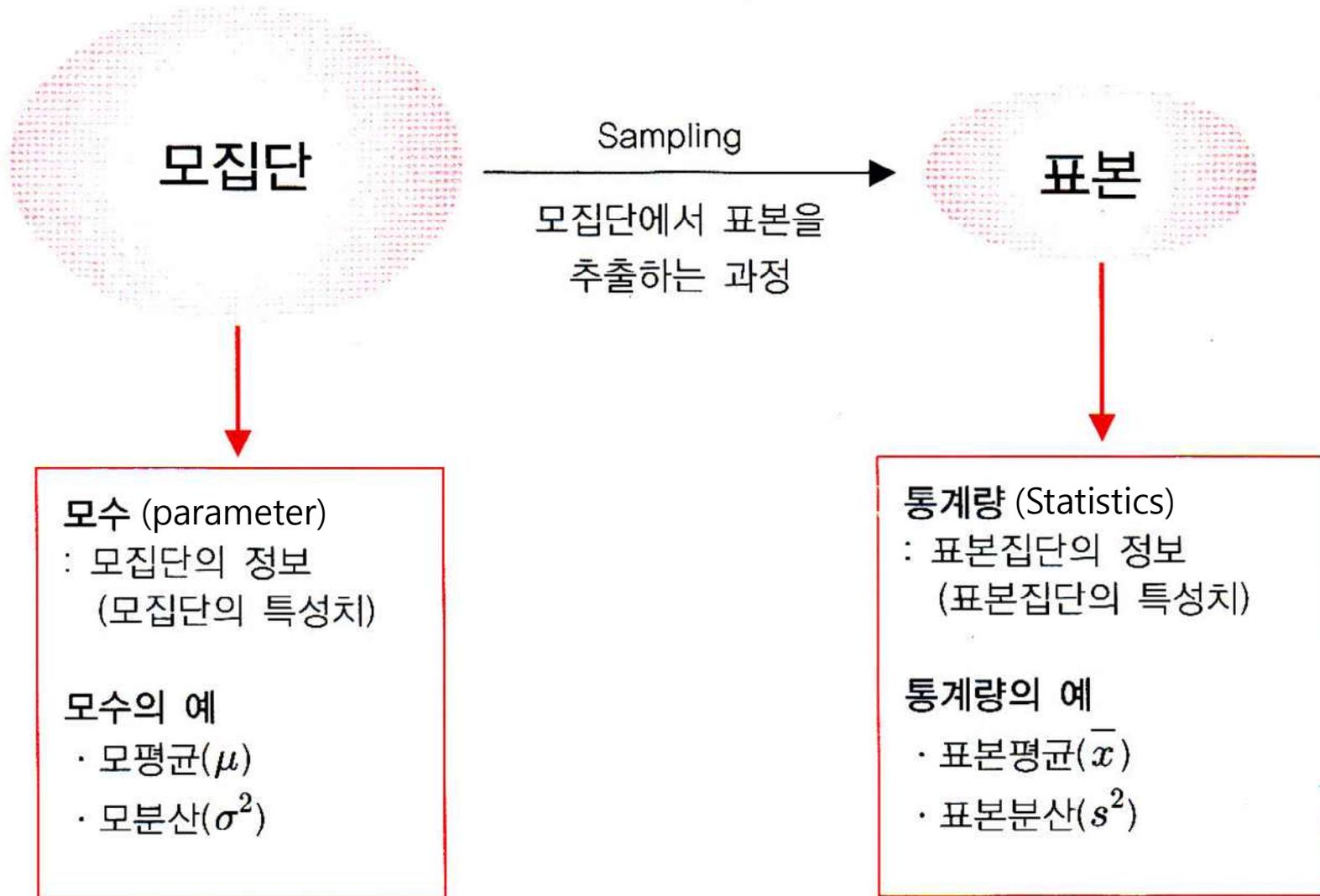
$\Rightarrow F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$  는 분자와 분모의 d.f.가  $m-1$  과  $n-1$  인 F 분포 따름

- ② 표기 :  $F \sim F(m-1, n-1)$

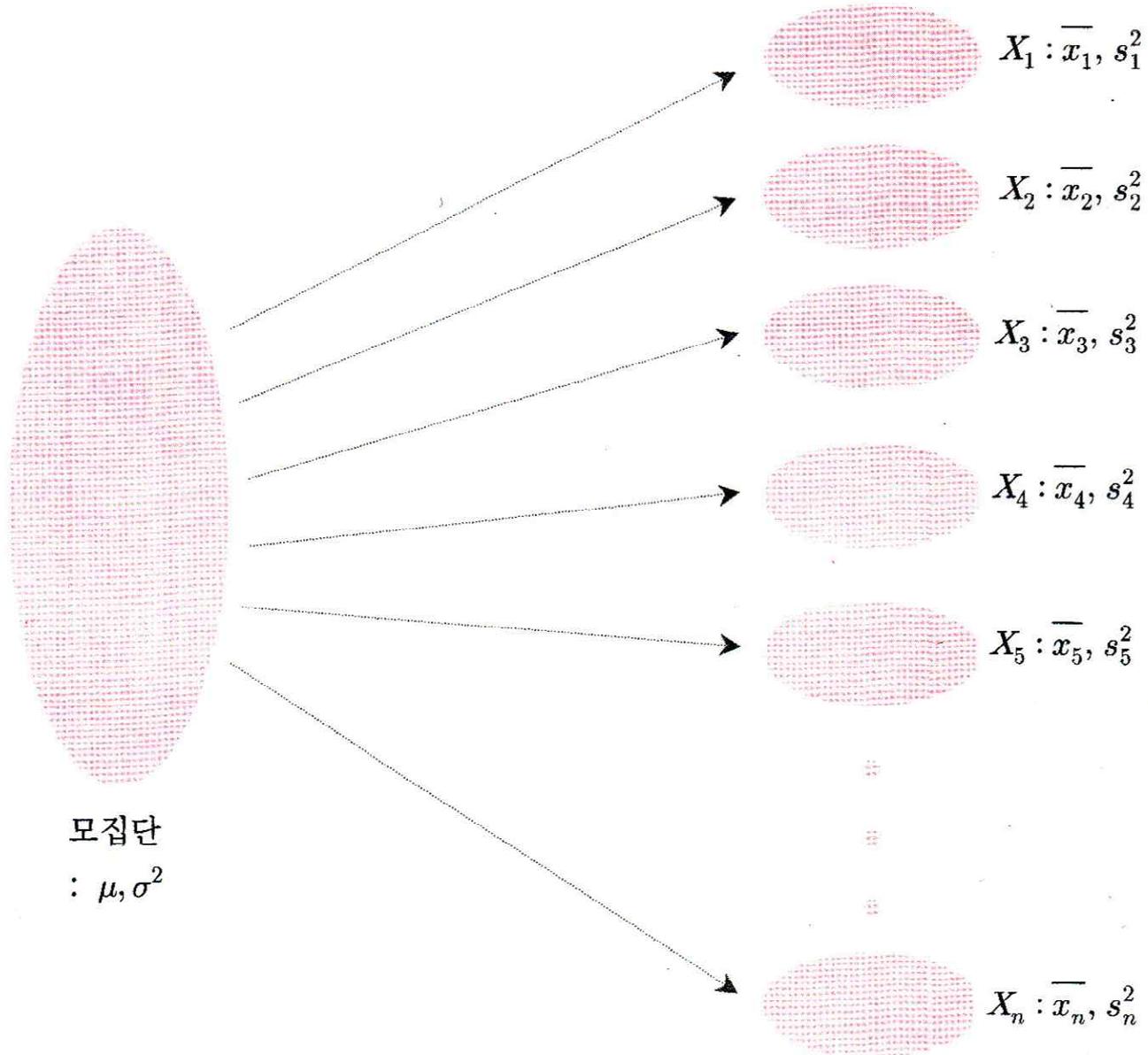
- ③ 분포의 형태:



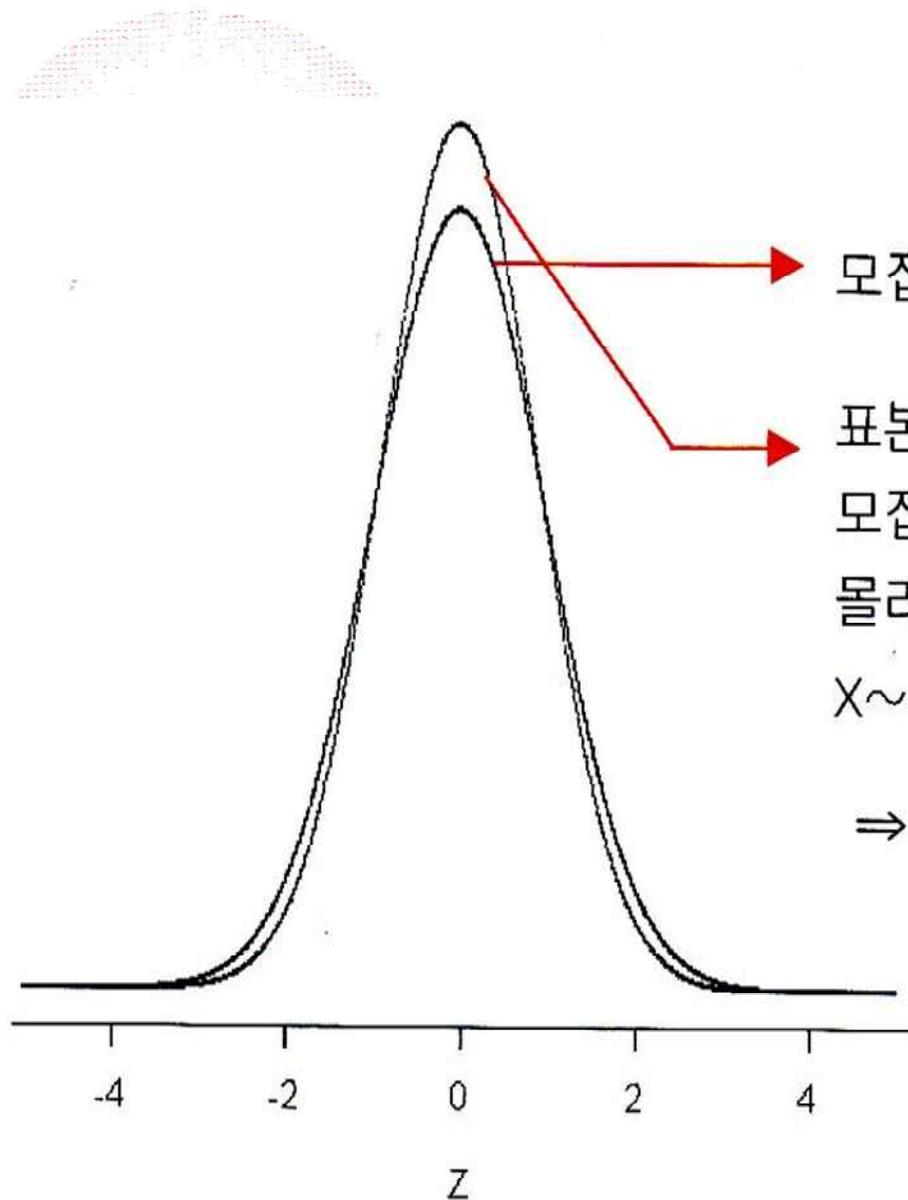
### 3. 표본분포 ( Sampling distribution)



# 표본분포 ( Sampling distribution )



# 표본분포 ( Sampling distribution )



모집단의 분포

표본평균의 분포

모집단보다 평균을 중심으로 많이 몰려 있음을 알 수 있다.

$$X \sim N(u, \sigma^2)$$

$$\Rightarrow \bar{X} \sim N\left(u, \frac{\sigma^2}{n}\right)$$

# 정규성 검정 (Normality test)

: 수집된 data들이 정규 분포의 가정을 만족하는지 검정하는 방법

< 방법 >

1) 먼저 자료의 분포 형태를 파악

2) 자료의 평균값과 중앙값의 차이가 많이 나지 않는지

: 중앙값 평균값이 비슷한지,  
5%절단평균 (5% Trimmed mean)

3) 자료의 왜도 (Skewness)와 첨도값 (Kurtosis)에 대한 확인

: -2~+2사이인지 ( 정규성을 만족한다는 가정하에 실시 )

4) Normal probability plot

: Normal probability plot과 비슷한 분포를 보이는지

5) Sapiro Wilks test/ Kolmogorov Smirnov test 실시

:  $P \geq 0.05$  인지

$H_0$  : 정규성을 만족

$H_1$  : 정규성을  
만족하지 않음

## 추정 (Estimation)

---

1. 모수적 (Parametric) 접근방법

: 주어진 자료로 "정규성 만족함" 을 알 수 있을 때

2. 비모수적 (Nonparametric) 접근방법

: 주어진 자료로 "정규성 만족하지 않음" 을 알 수 있을 때

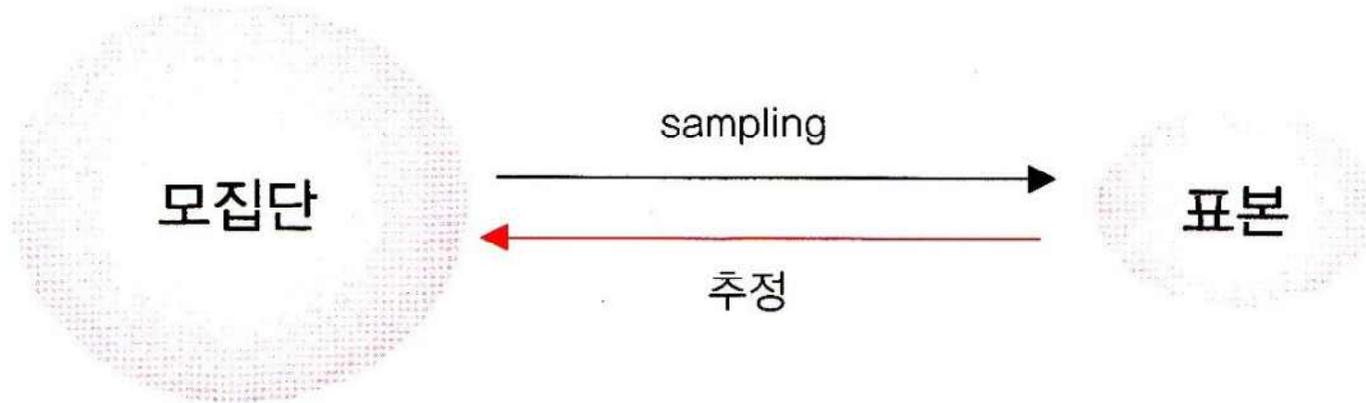


1) 하나의 모집단인 경우

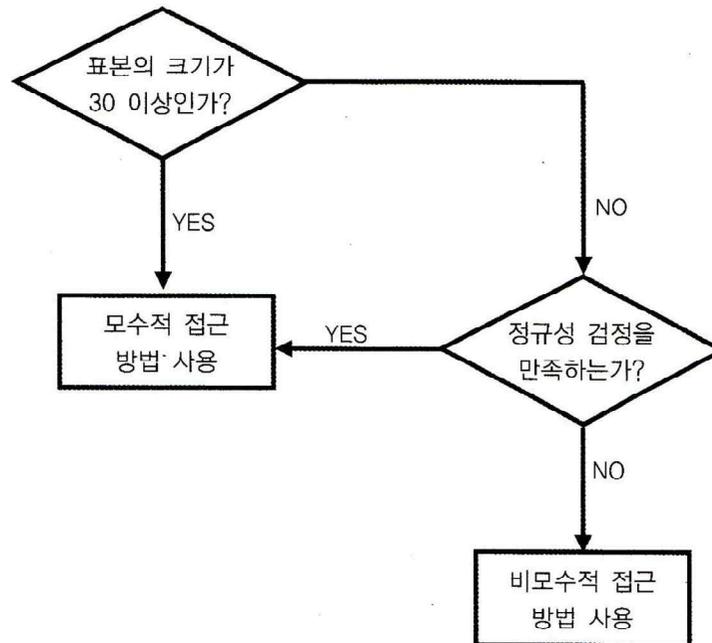
2) 두 개의 모집단인 경우

3) 하나의 모집단 전, 후 비교의 경우

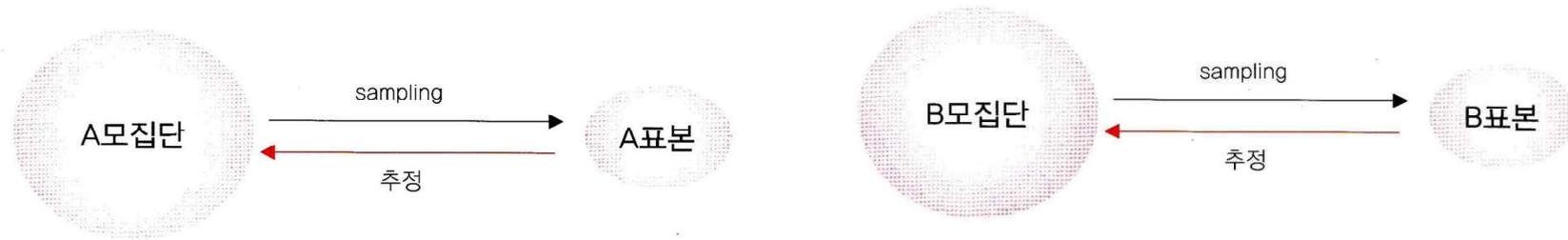
# 1) 하나의 모집단인 경우



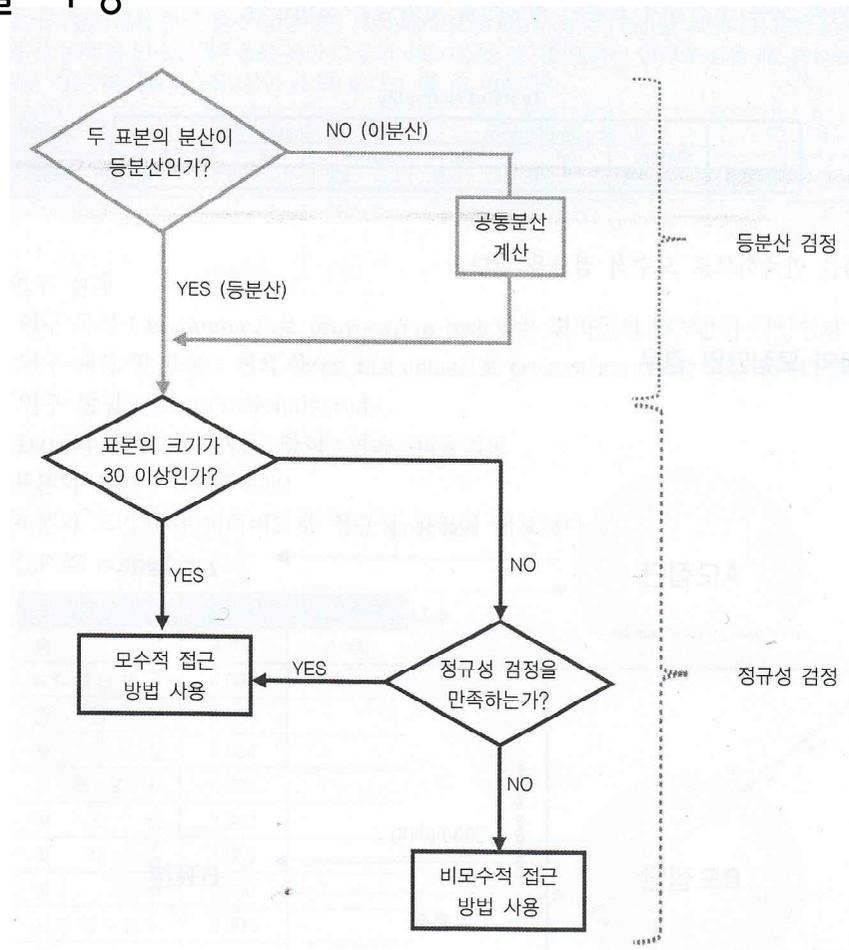
- 추정 전 확인해야 할 사항



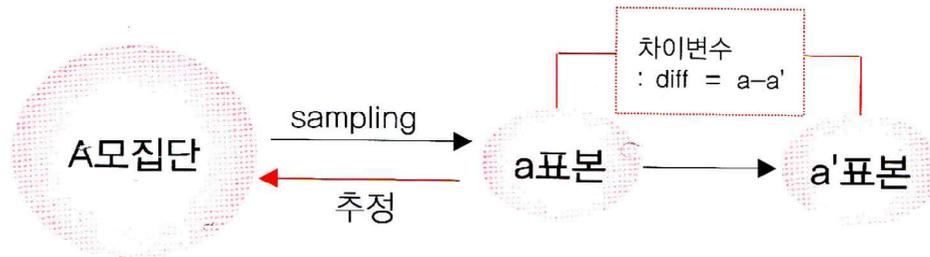
## 2) 두 개의 모집단인 경우



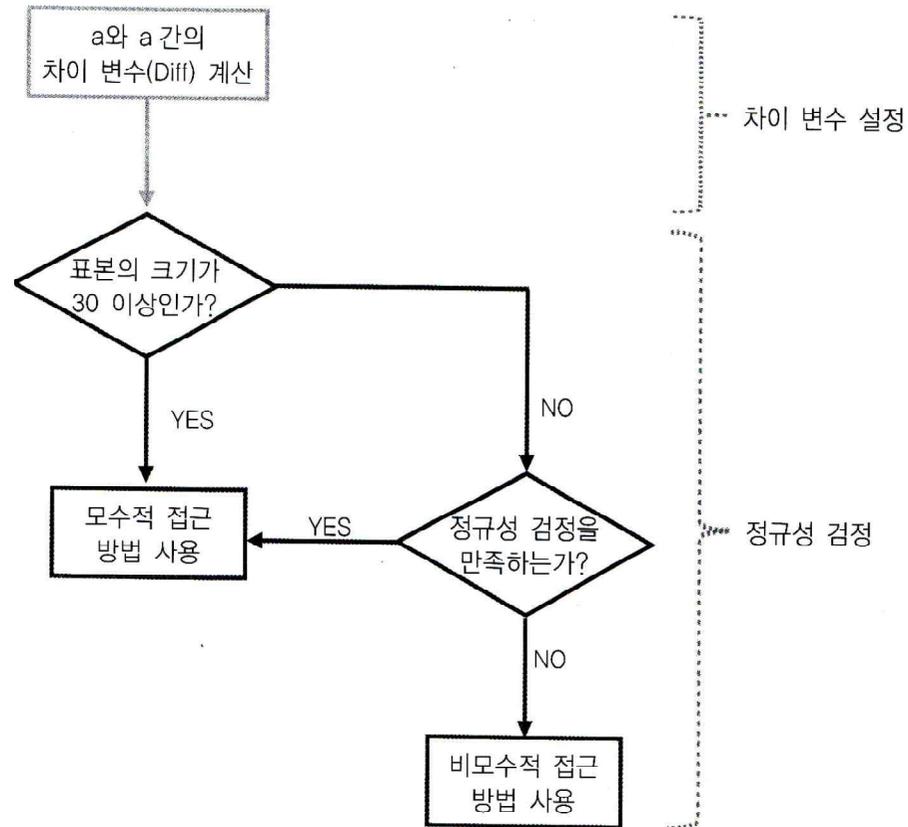
- 추정 전 확인해야 할 사항



### 3) 하나의 모집단이며, 종속적 관계인 경우



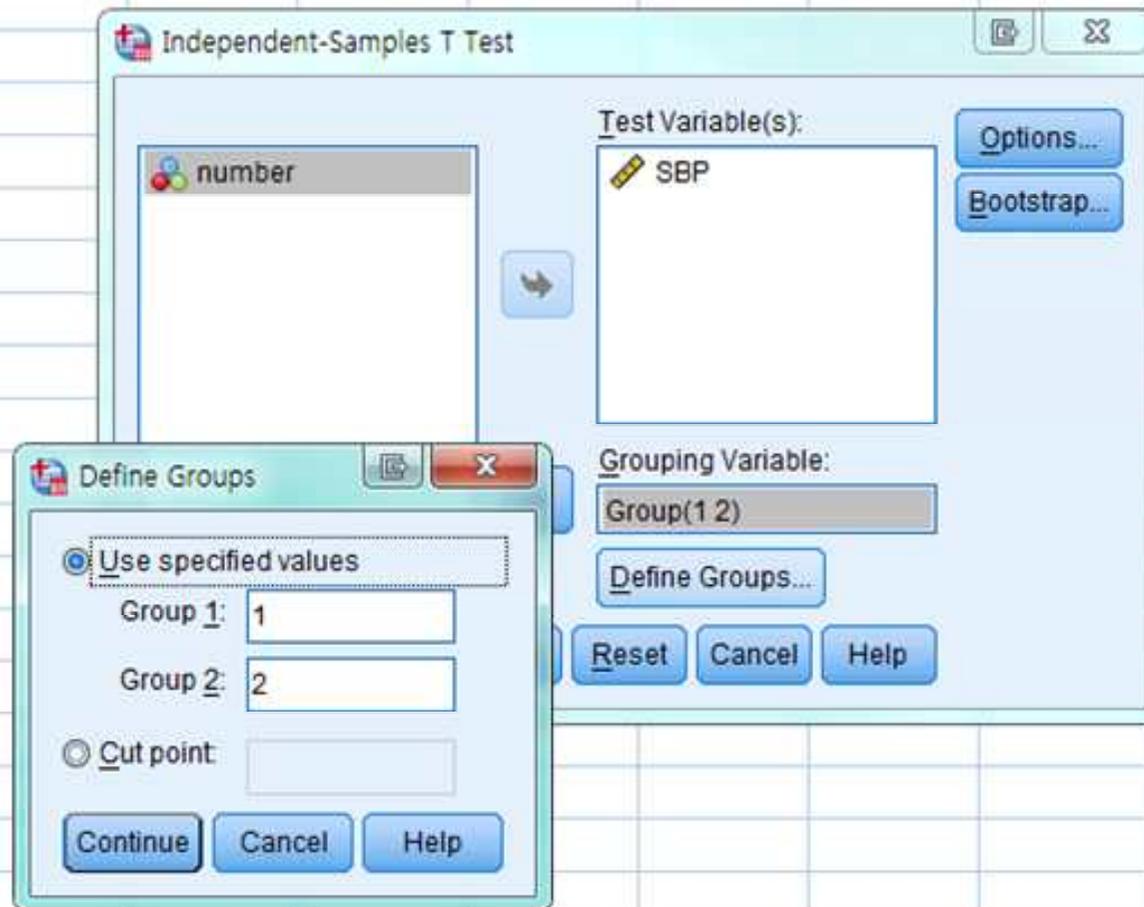
- 추정 전 확인해야 할 사항



number	Group	SBP
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	1	
10	1	
11	2	
12	2	
13	2	
14	2	
15	2	
16	2	
17	2	
18	2	
19	2	
20	2	

The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and the 'Compare Means' option is selected. The sub-menu for 'Compare Means' is also open, showing options like 'Means...', 'One-Sample T Test...', 'Independent-Samples T Test...', 'Paired-Samples T Test...', and 'One-Way ANOVA...'. The 'Independent-Samples T Test...' option is highlighted. The data table in the background has a red horizontal line under row 10.



Group Statistics

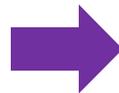
Group	N	Mean	Std. Deviation	Std. Error Mean
SBP 1	10	144.80	3.553	1.123
SBP 2	10	140.60	5.661	1.790

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SBP	Equal variances assumed	2.067	.168	1.987	18	.062	4.200	2.113	-.240	8.640
	Equal variances not assumed			1.987	15.138	.065	4.200	2.113	-.301	8.701

- Levene's test for equality of variance : 등분산 검정

p-value > 0.05 : 등분산  
 p-value < 0.05 : 이분산



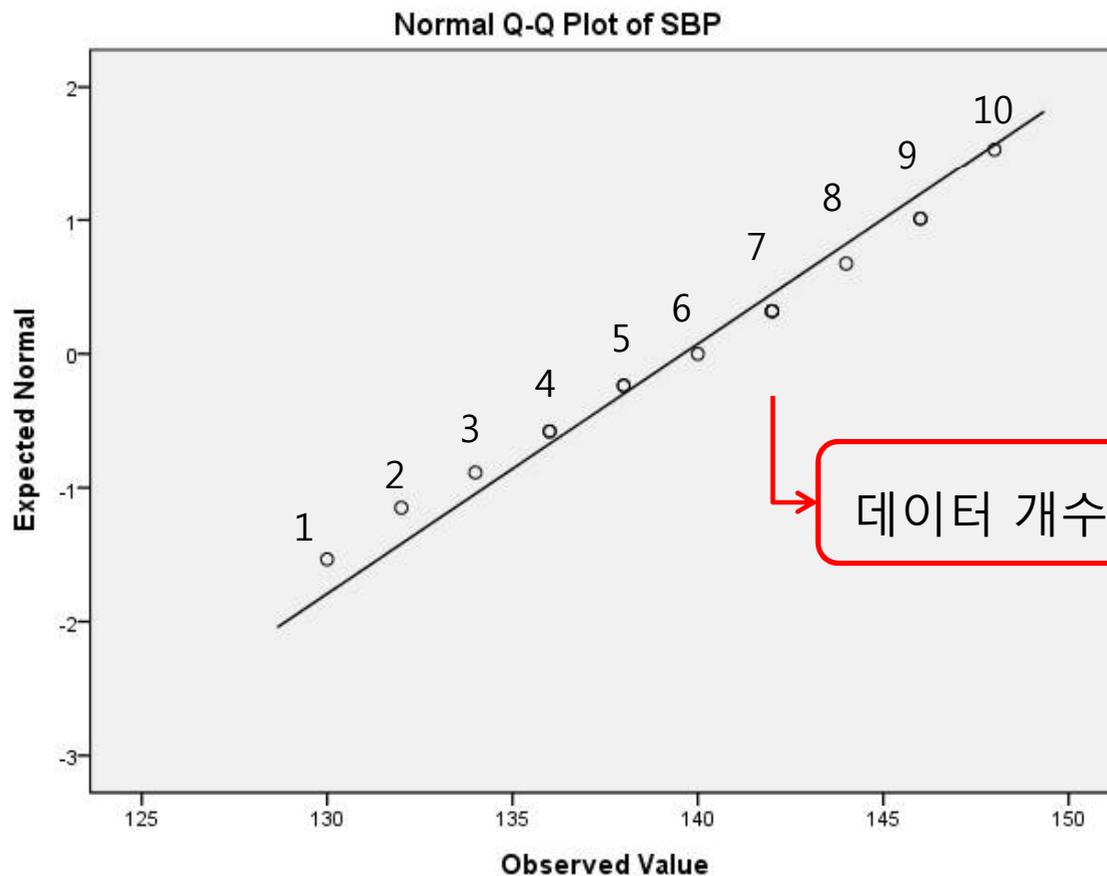
**P = 0.168 : 등분산**

## 2. Q-Q (Quantile-quantile) plot

- Quantile : 모집단(population) 또는 표본의 전체도수(sample frequency)를  
오름차순으로 정렬 - > n 등분을 함
  - percentile : 백분위수
  - quartile : 4 분위수
  - quintile : 5 분위수

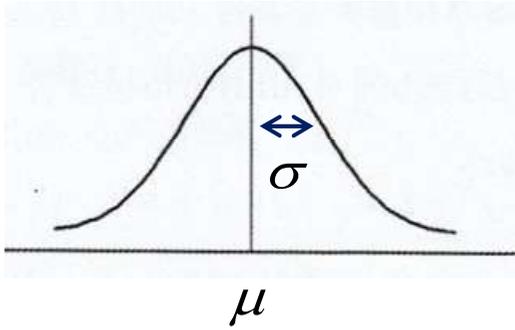
어느 A 대학병원에서 3개월간 ACEi로 치료받고 있는 환자 50명중 15명(SRS으로 추출)의 수축기 혈압(mmHg)을 재어 보았더니 같았다. 3개월간 고혈압 치료를 받은 환자들의 평균혈압은 140보다 작다고 말할 수 있는가?

index	1	2	3	4	5	6	7	
SBP	148	134	138	142	142	142	146	
index	8	9	10	11	12	13	14	15
SBP	146	130	138	140	136	132	144	136

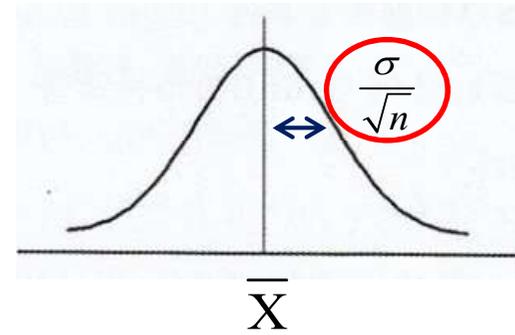


노란 동그라미 : 10개

데이터 개수 : 10개



**Sampling** →



Standard error ( $\bar{X}$ 의 표준편차)

$$= \frac{\sigma}{\sqrt{n}}$$

