

## Chapter04

## 통계분석 II

# I 분산분석 (ANOVA)

M T W T F S S

FASTCAMPUS  
ONLINE

금융공학/퀀트 I

강사. 장순용

# I 키포인트

- 분산분석 (ANOVA).
- 일원 분산분석.
- 이원 분산분석.



## I 분산분석

- 세개 이상의 모집단 (표본)이 있는경우 평균을 비교 검정하고자 한다.
- 표본들을 짝지어서 t 검정을 하는 것은 번거롭고 오류가 누적될 수 있다.

⇒ 단 한번에 다음과 같은 가설을 검정하고자 한다.

“귀무가설”  $H_0: \mu_1 = \mu_2 = \mu_3 = \dots$

“대립가설”  $H_1: \mu_i \neq \mu_j$ 와 같은 경우가 있다.

## I 분산분석

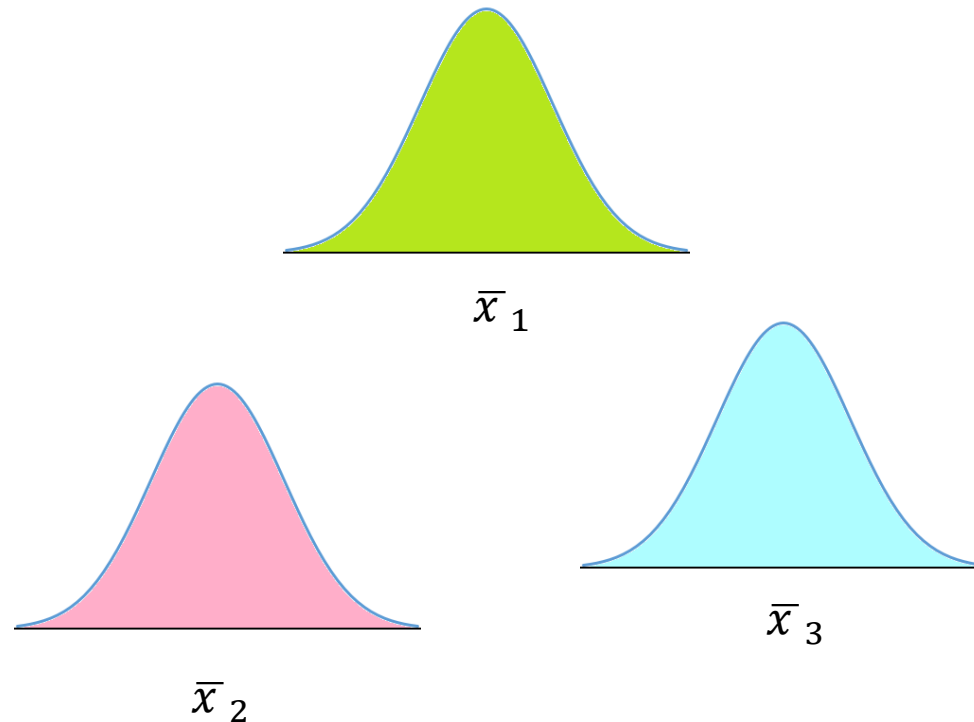
- 일원 분산분석 : **한개**의 수치형 변수  $Y$ 와 **한개**의 명목형 변수  $X$ .  
 →  $Y$ 는 종속변수이고  $X$ 는 독립변수이다.  
 예). 여러 반  $X$ 의 성적  $Y$ 를 비교 검정한다.
- 이원 분산분석 : **한개**의 수치형 변수  $Y$ 와 **두개**의 명목형 변수  $X_A$ ,  $X_B$ .  
 →  $Y$ 는 종속변수이고  $X_A$  와  $X_B$ 는 독립변수이다.  
 예). 반  $X_A$ 과 성별  $X_B$ 로 그룹을 나누어서 성적  $Y$ 를 비교 검정한

## I 분산분석

- 다음과 같은 조건을 전제한다:
  - 종속변수  $Y$ 는 정규분포를 따른다.
  - 표본이 대표하는 모집단의 분산은 동일하다.
  - 표본은 독립적으로 표집되었다.

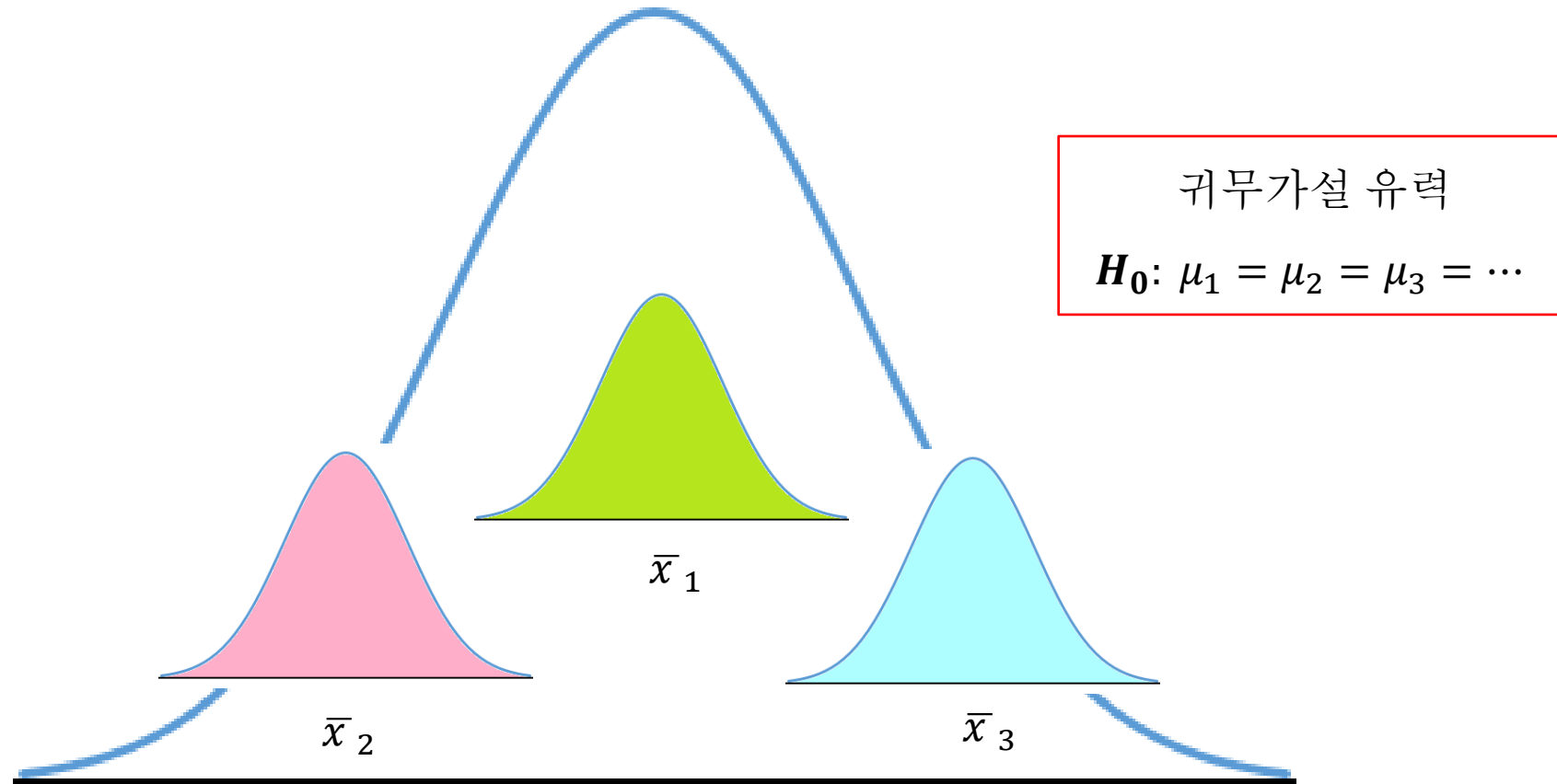
# I 일원 분산분석

- 다음과 같이 여러개의 표본 (처리)가 있을때 이들의 표본평균을 비교해 본다.  
→ 모집단의 평균은 서로 같은가?



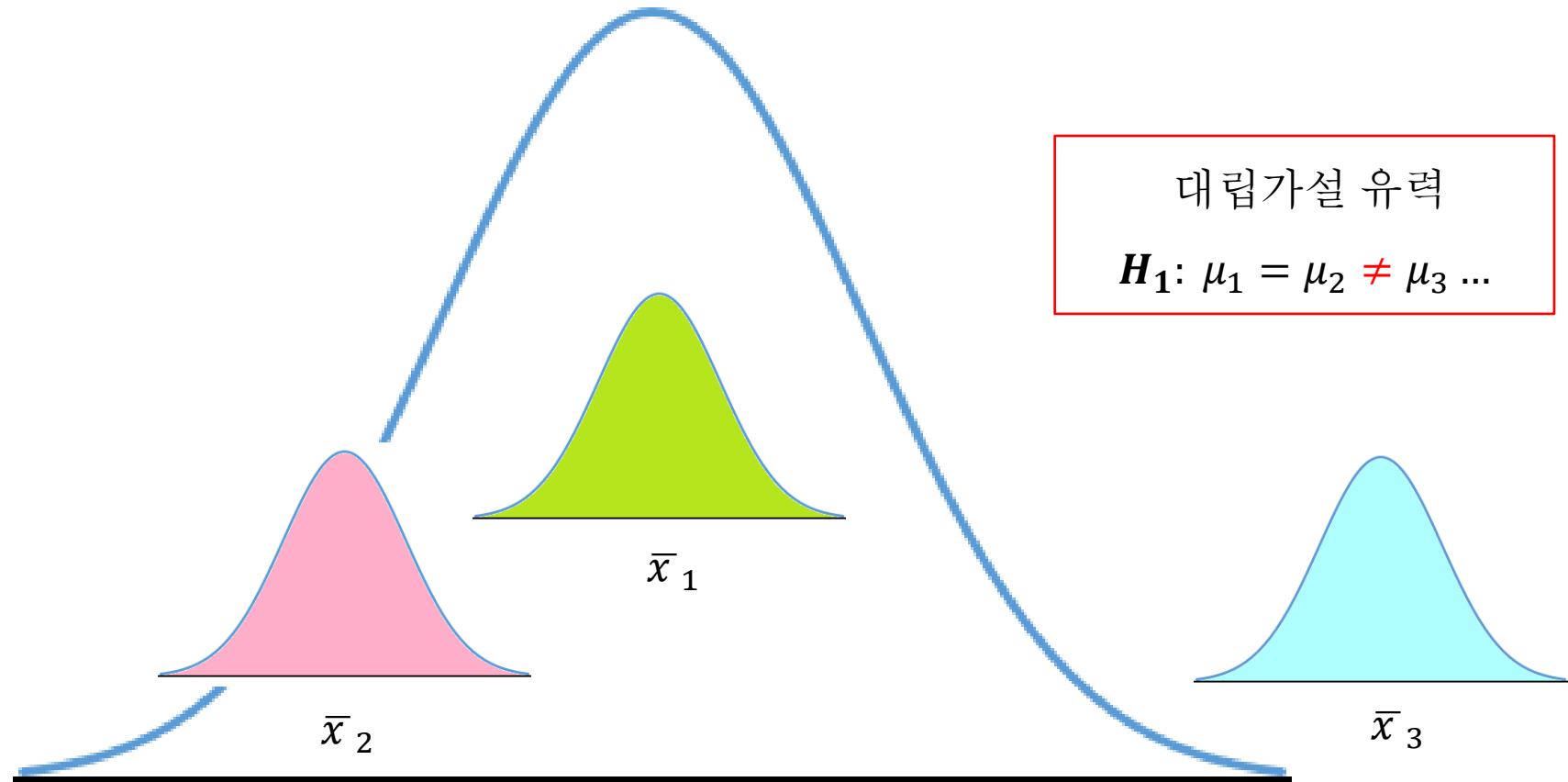
# I 일원 분산분석

- 전체적인 데이터 분포와 비교했을 때 표본평균들이 아래와 같이 모여 있을 수 있다.



# I 일원 분산분석

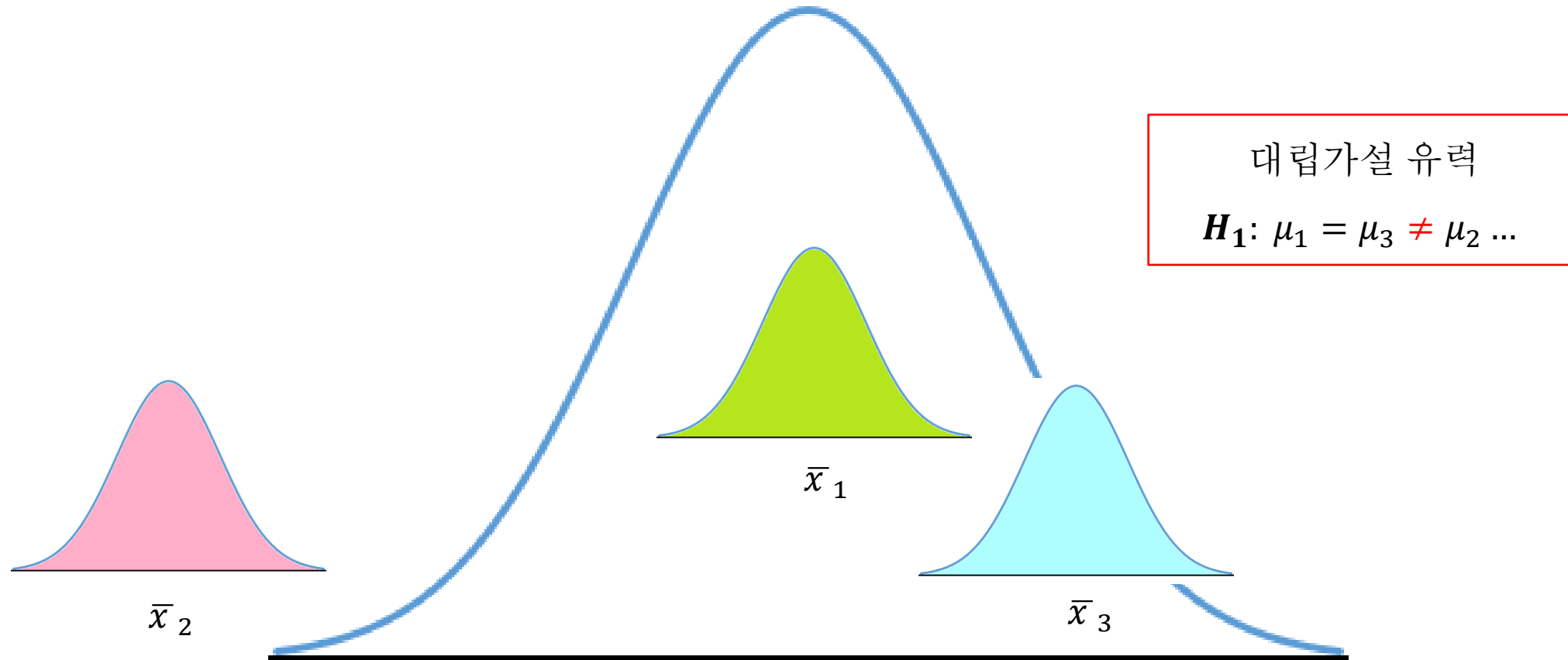
- 아니면 몇 개의 표본평균이 나머지와 많이 떨어져 있을 수도 있다.





# I 일원 분산분석

- 아니면 몇 개의 표본평균이 나머지와 많이 떨어져 있을 수도 있다.



# I 일원 분산분석

- 다음과 같이 검정통계량을 만들어 본다:

$$\text{검정통계량} = \frac{\text{표본평균 사이의 분산 (between)}}{\text{표본내의 분산 (within)}} = \frac{MStr}{MSE}$$

$$\rightarrow \text{검정통계량} = \frac{\text{큼}}{\text{작음}} \Rightarrow H_0 \text{ 기각 후 } H_1 \text{ 채택.}$$

$$\rightarrow \text{검정통계량} = \frac{\text{비슷}}{\text{비슷}} \Rightarrow H_0 \text{ 유지. 조금은 애매한 경우.}$$

$$\rightarrow \text{검정통계량} = \frac{\text{작음}}{\text{큼}} \Rightarrow H_0 \text{ 유지. 모든 표본평균이 가까이 모여있음.}$$

# I 일원 분산분석

- 다음과 같이 검정통계량을 만들어 본다:

$$\text{검정통계량} = \frac{\text{표본평균 사이의 분산 (between)}}{\text{표본내의 분산 (within)}} = \frac{MStr}{MSE}$$

- 분산의 비율이므로 F확률분포를 따른다 ⇐ 우측검정.
- 분자의 자유도 = 표본(처리)의 개수 - 1.
- 분모의 자유도 = 전체 데이터의 개수 - 표본(처리)의 개수.

## 10 이원 분산분석

- 이원 분산분석 : **한개**의 수치형 변수  $Y$ 와 **두개**의 명목형 변수  $X_A$ ,  $X_B$ .

**예**). 반  $X_A$ 과 성별  $X_B$ 로 그룹을 나누어서 성적  $Y$ 를 비교 검토한  
다.

- $X_A$ 에는  $p$ 개의 처리가 있고  $X_B$ 에는  $q$ 개의 처리가 있다고 가정한다.

## I 01원 분산분석

- 귀무가설과 대립가설은 각각  $A$ 와  $B$ 로 구분한다.

⇒ “귀무가설”  $H_0: \mu(A_1) = \mu(A_2) = \dots = \mu(A_p)$

“대립가설”  $H_1$ : 적어도 하나의 모평균은 나머지와 다르다.

⇒ “귀무가설”  $H_0: \mu(B_1) = \mu(B_2) = \dots = \mu(B_q)$

“대립가설”  $H_1$ : 적어도 하나의 모평균은 나머지와 다르다.

## 10이원 분산분석

- 마찬가지로 검정통계량도 각각  $A$ 와  $B$ 로 구분되며 F분포를 따른다.

$$\text{검정통계량}_A = \frac{MStr_A}{MSE} \sim F(p-1, (p-1) \cdot (q-1))$$

$$\text{검정통계량}_B = \frac{MStr_B}{MSE} \sim F(q-1, (p-1) \cdot (q-1))$$



# I O원 분산분석

- 이원 분산분석은 without replication (반복 없음)과 with replication(반복 있음)으로 나뉜다:

→ 반복 없음: 유형별 단 한개의 자료값이 있음.

aov()에서  $Y \sim X_A + X_B$ 와 같은 수식을 사용.

→ 반복 있음: 유형별 여러개의 자료값이 있음.

aov()에서  $Y \sim X_A * X_B$ 와 같은 수식을 사용하여 상호작용 포함.

I 끝.

감사합니다.

