◎ Two sample test

⇒ Pooled variance 끼리 가능 계산 과정을 이해하기 쉽도.

◎ 등분산 가정 (Equality of variance)

① 정규성 가정 test
② 등분산 가정 test ← Sp 사용하기 때문에
③ t-test

if 등분산 O → t-test
등분산 X → welch's t-test

◎ 등분산 검정

leven Test → p-value 해석 → 5% 미만으로 떨어지면 귀무가설 기각.

기본 가설 $H_1 : \sigma_1^2 \neq \sigma_2^2$

$H_0 : \sigma_1^2 = \sigma_2^2$

* 기본적으로 welch's test 한다.
∵ 등분산 유무 상관없이 사용 가능.

◎ 일원배치 분산분석 (one-way ANOVA)
$\underbrace{\text{ANOVA}}_{\text{Analysis of Variance.}}$

· 요인 factor 가 하나이고, 수준수 level or treatment 가 3개 이상일 경우에 수준간 자료가 통계적 서로 다른 모집단으로 부터 생성되었는지 검증.

비교

ctrl, trt1, trt2, ...

← 이렇게 여러번 검증하면 error가 증가함 (e.g. 0.05 + ...)

요인 ++

비교 / 요인	Ctrl	trt1	trt2
항목			
구형도			

이원배치
↓
다원배치

④ 일원배치 분산분석

- 분산분석의 가정

각각 평균은 다르다.

$$X_{ij} \overset{iid}{\sim} \boxed{N}(\mu_i, \sigma^2), \quad \forall i, j$$

분산은 같다.

i : 그룹

j : data

X_{ij} $i = 1, 2, \ldots, b$

$j = 1, 2, \ldots, n_i$

- 독립성, 정규성, 등분산성

가정을 분자를 대체할 수 있음.

$p < 0.05$ (유의확률)

$\Downarrow$

"유의하다"

- 가설

H_0 : Not H_1 : $\mu_1 = \mu_2 = \cdots = \mu_b$

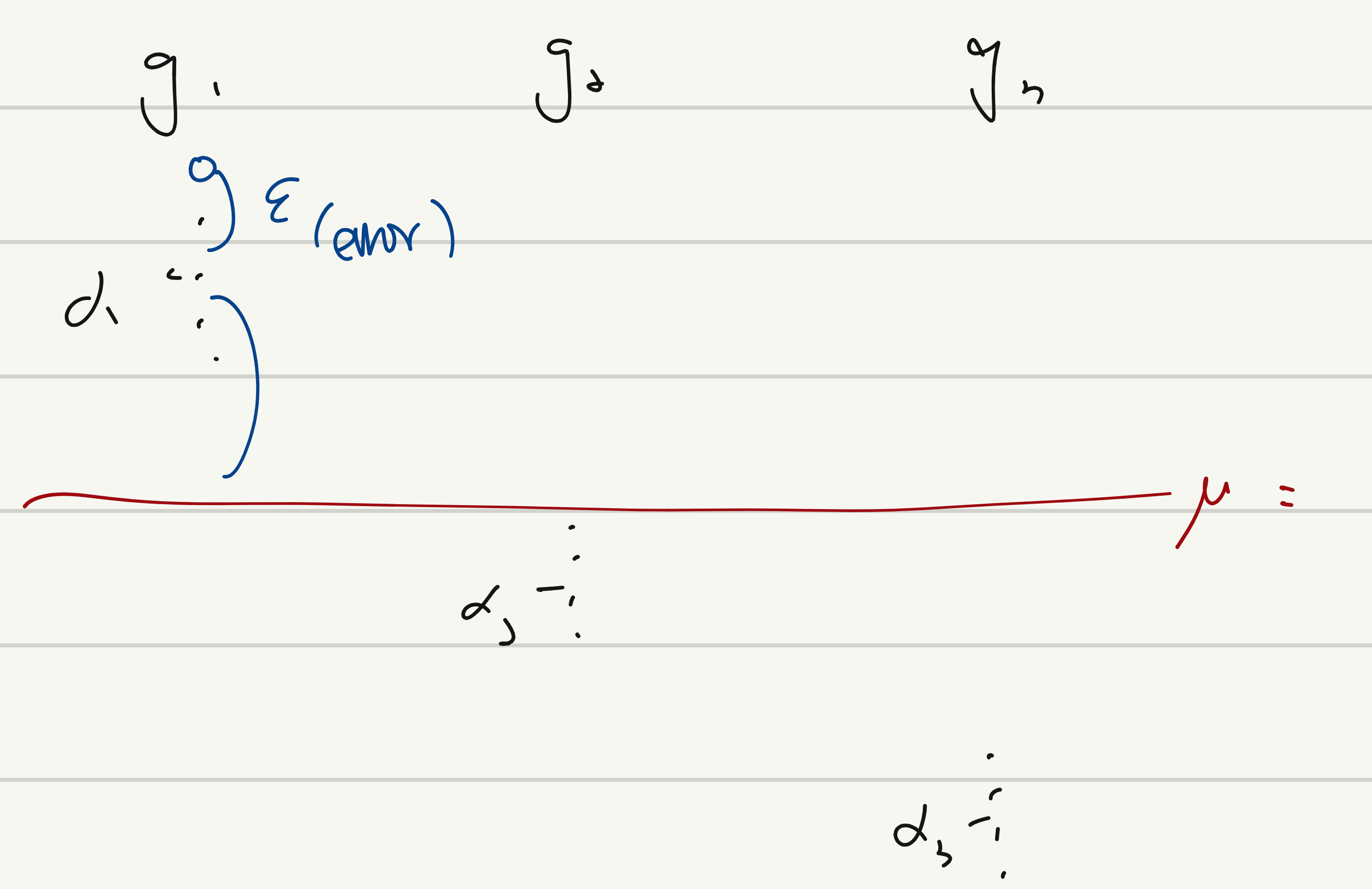
H_1 : 그룹간 평균에 차이가 있다

⑤ ANOVA 구조모형

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \quad \forall i, j$$

ex



$$\mu = \frac{\sum x_{ij}}{N} = \bar{x}$$

g_1 g_2 g_3

μ : 전체 a 평균

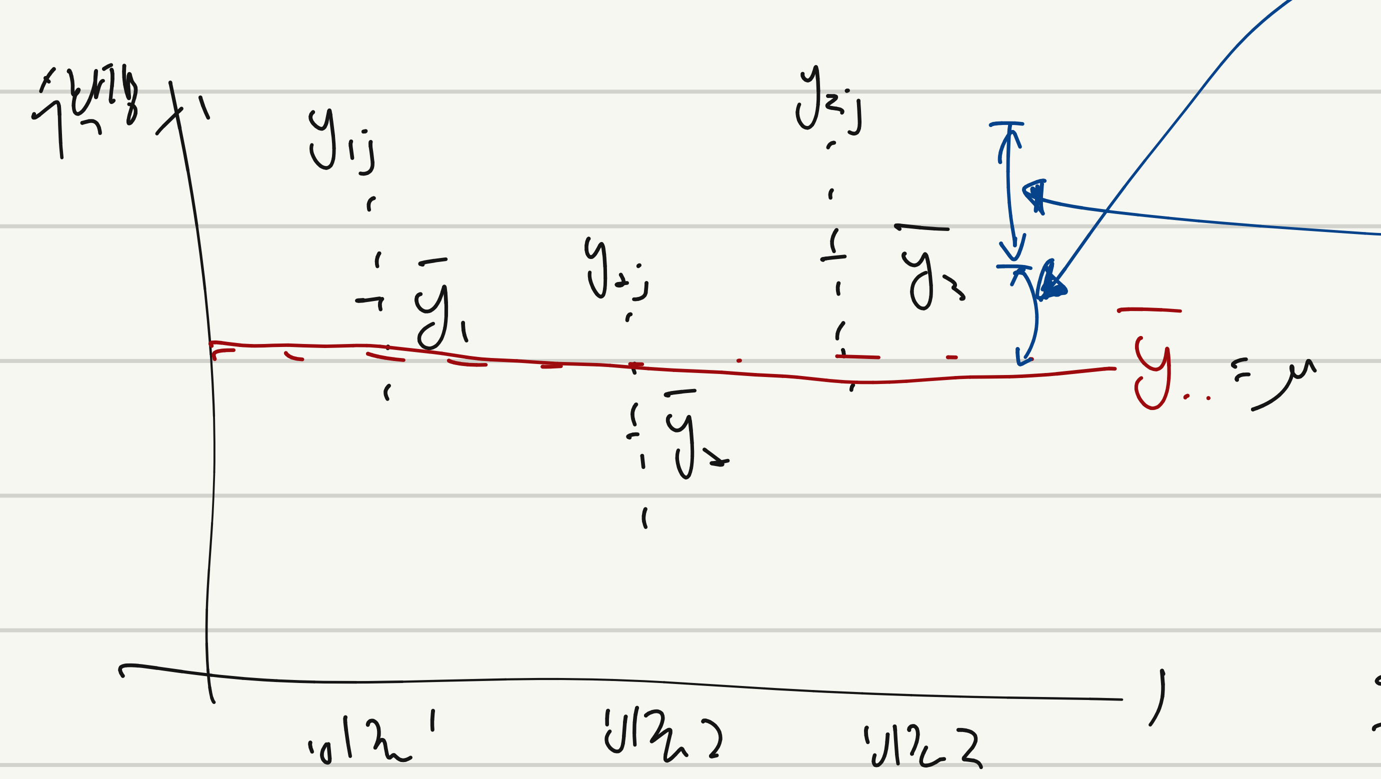
$\mu_1 -$

α_i : treatment에 의한 효과

(비교 μ_2, 2, 3)

$$\sum_{\mu_3} \alpha_i = 0$$

⑥ 총변동의 분할



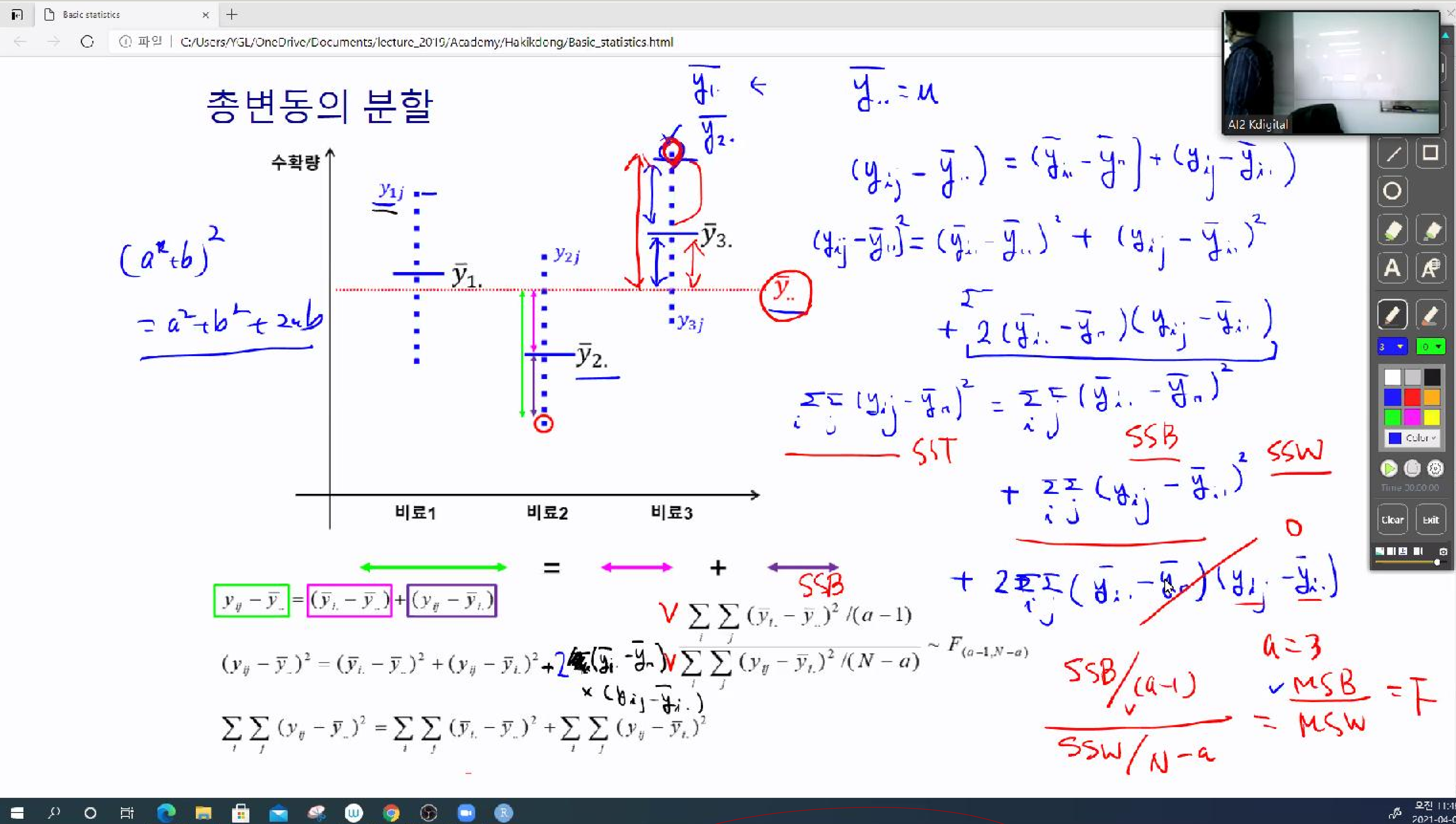
$$y_{ij} - \bar{y} = (\bar{y}_{i\cdot} - \bar{y}) + (y_{ij} - \bar{y}_{i\cdot})$$

$$(y_{ij} - \bar{y})^2 = (\bar{y}_{i\cdot} - \bar{y})^2 + (y_{ij} - \bar{y}_{i\cdot})^2 + 2(\bar{y}_{i\cdot} - \bar{y})(y_{ij} - \bar{y})$$

$$\underbrace{\sum_i \sum_j (y_{ij} - \bar{y})^2}_{SST} = \sum\sum \underbrace{A}_{SSB} + \sum \underbrace{B}_{SSW} + \cancel{2\sum\sum AB} = 0$$

Sum of Square

- Total
- Between
- Within



그룹간 분산 ≈ 1
그룹내 분산

F value는 항상 양수.
$F(a-1,\ n-a)$

F-value가 "2"보다 커지면 "유의하다"

$H_0:\ \mu_1 = \mu_2 = \mu_3$

⑥ 분산분석표의 작성. (ANOVA Table)

Source of variation	Sum of squares (SS)	Degrees of freedom (DF)	Mean Square (MS)	F-statistic
Treatments	$SS_{between}$ (SS_b)	k-1	$MS_b = SS_b/(k-1)$	$F = MS_b/MS_w$
Error (or Residual)	SS_{Within} (SS_W)	N-k	$MS_W = SS_W/(N-k)$	
Total	SS_{Total} (SS_T)	N-1		

https://sixsigmastudyguide.com/anova-analysis-of-variation

⑥ 사후 검정. Post_hoc

Multiple comparison problem.

FWER (Family wise type 1 error rate)

$$= 1 - (1-\alpha)^n \le n\alpha$$

if $\alpha = 0.05$, $n = 12$

$$\Rightarrow 1 - (0.95)^{12} = 0.45 \rightarrow 45\%$$

100번중에 50번은 우연하게 나온다.

2번 중 1번은 기각될 수 있음 ⇐ 이러면 안됨 ⇐

⑥ FWER correction

- Bonferroni : $\underbrace{\text{Set } \alpha_c = \frac{\alpha}{n}}$ 이 값보다 p-value 작으면

  ⇒ 이 값이 p adj

- Tukey : 각 수준상 표본의 개수가 동일

- Holm - Bonferroni ← 많이 사용.

  ⋮

  여러가지 있음.

* ANOVA 에서 유의 해야지만, post hoc (사후검정.) 이 의미가 있음.

$$\underbrace{\text{normality test} \rightarrow \text{equality of variance} \rightarrow \text{ANOVA} \rightarrow \text{Post hoc}}_{\text{기본 틀.}}$$

⑥ 이원 배치 분산분석 (two-way ANOVA)

사료 요인1

	A	B	C
품종 요인2 Kor			
USA			
Jap			

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

비료 품종

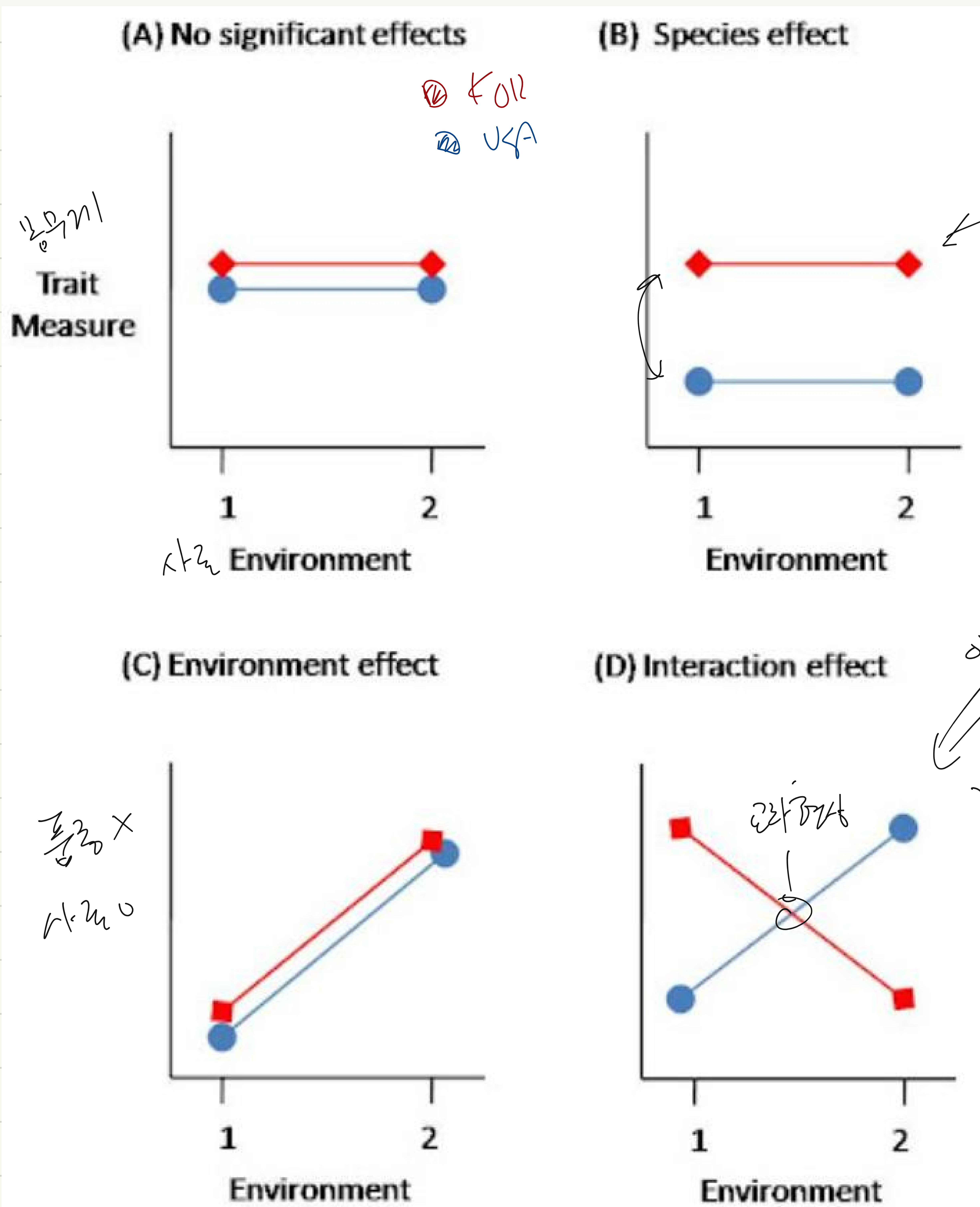
$$\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2), \ \forall i, j, k$$

사료 품종 개수

※ 상호작용 효과 (Interaction effect)

① Kor
② USA

몸무게



사료

품종간의 effect만 있다

품종 X
사료 O

교차점

이건 원래는 ANOVA로 검증할 수 없다.

몸무게

품종 O
사료 O