

7-1

Repeated Measure ANOVA란?

- 반복 측정 분산분석 (1) -

* 개념: 전치나 집단에 상대하여 하나하나의 날개를 이룬 것
하나의 독립된 생물체

Repeated Measure ANOVA

동일한 개체에 대한

각 개체들이 대하는 동일한

- 반복 측정 분산분석 (Repeated Measure ANOVA)란?

측정을 반복적으로

- 말 그대로 반복 측정된 데이터를 이용한 분산분석임

시행함.)

- 예제
 - 투약효과 확인을 위해 남녀 25명씩 50명 4회 콜레스테롤 레벨 측정
 - 운동효과 측정을 위해 남녀 3명씩 6명 3회 운동레벨 측정

⌚ Time point 1

⌚ Time point 2

⌚ Time point 3



반복 측정 분산분석 (Repeated Measure ANOVA)

X 측정을 두 번하는 것이라면, Paired sample T-test
 측정을 세 번 이상 하는 것이라면, RM ANOVA.

- 예제
 - 운동효과 측정을 위해 남녀 3명씩 6명 3회 운동레벨 측정

X. 동일변수: 측정시기
 종속변수: 운동레벨

X. RM로 당연히
 동일변수는
 '영역형'이고
 종속변수는
 '연속형'이다.

| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

증속변수가 세 개 인것이!!!

반복 측정 분산분석 (Repeated Measure ANOVA)

- 만약 One-way ANOVA를 적용할 경우
 - Within variance가 비정상적으로 커짐

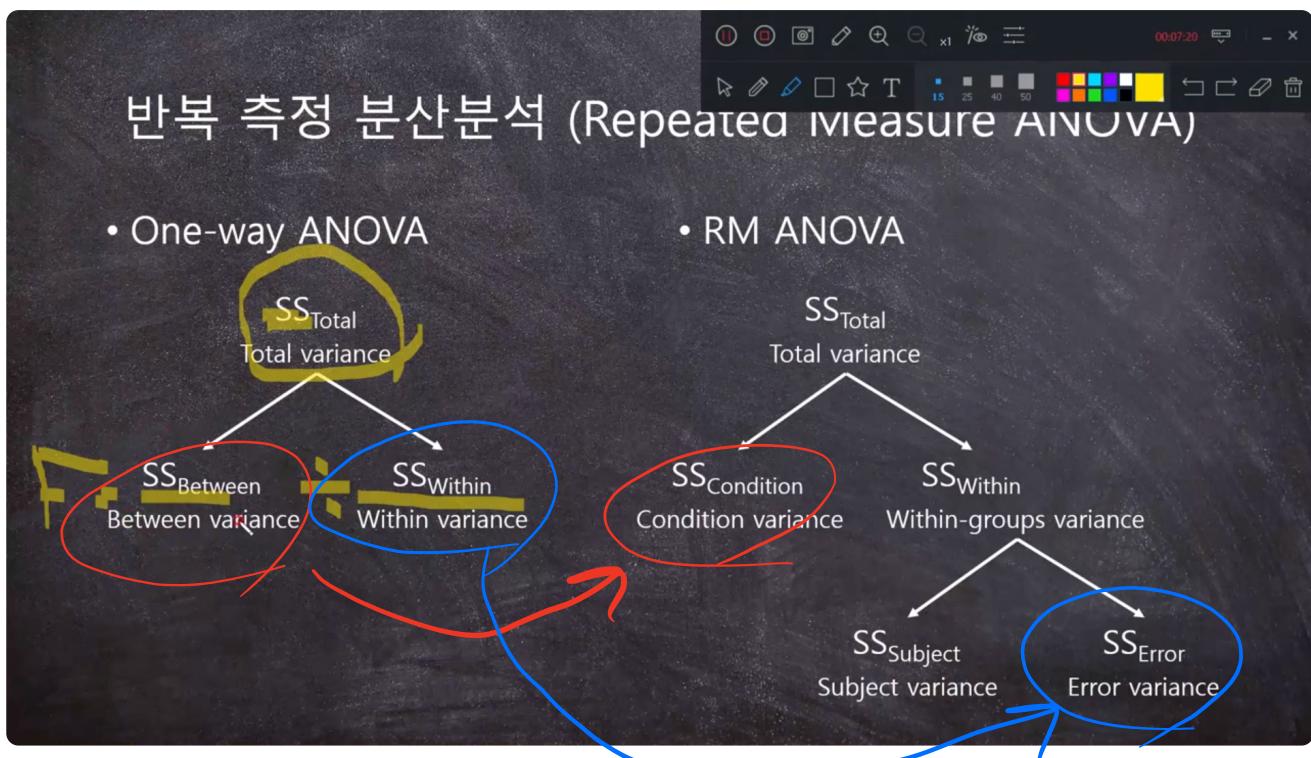
| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

Within variance

correlation

반복 측정 분산분석 (Repeated Measure ANOVA)

- 그렇다고 각 사람을 그룹으로 보고 two-way ANOVA를 하면
 - 그룹별로 값이 단 한 개만 존재
 - within variance를 구할 수 없음
- 결론적으로 기존의 방법으로 접근할 경우
 - within variance가 비정상적으로 증가
 - F-value(F값)이 감소
 - $F\text{-value} = \text{between variance} \div \text{within variance}$
 - 따라서, 결론적으로 p-value가 작아져 유의해야할 값이 유의하지 않음



반복 측정 분산분석 (Repeated Measure ANOVA)

• RM ANOVA

- SS_{Time} (time variance) *condition variance (= Between variance)*

$$\begin{aligned}
 &= 6 \times (42.8 - 45.9)^2 \\
 &+ 6 \times (45.3 - 45.9)^2 \\
 &+ 6 \times (49.7 - 45.9)^2 \\
 &= 143.44
 \end{aligned}$$

- df of SS_{Time}
 $= k - 1$
 $= 3 - 1 = 2$ (k는 time의 개수)

| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|---------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

반복 측정 분산분석 (Repeated Measure ANOVA)

• RM ANOVA

- $SS_{Within-groups}$ (within-groups variance)

$$\begin{aligned}
 &= (45 - 42.8)^2 + (42 - 42.8)^2 \\
 &+ (36 - 42.8)^2 + (39 - 42.8)^2 \\
 &+ \\
 &+ (59 - 49.7)^2 + (56 - 49.7)^2 \\
 &= 715.5
 \end{aligned}$$

- df of $SS_{Within-groups}$
 $= N - k$
 $= 18 - 3 = 15$ (N은 전체 샘플 개수, k는 time의 개수)

| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|---------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

반복 측정 분산분석 (Repeated Measure ANOVA)

- RM ANOVA
 - SS_{Subject} (subject variance)

$$= 3 \times (50.0 - 45.9)^2$$

$$+ 3 \times (43.0 - 45.9)^2$$

$$+ \dots$$

$$+ 3 \times (49.7 - 45.9)^2$$

$$= 658.3$$
 - df of SS_{Subject}

$$= n - 1$$

$$= 6 - 1 = 5$$
 (n은 time 그룹내 샘플 개수)

| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|---------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

반복 측정 분산분석 (Repeated Measure ANOVA)

이가 진짜 Within variance가 된다!

- RM ANOVA
 - SS_{Error} (error variance)

$$= SS_{\text{Within-groups}} - SS_{\text{Subject}}$$

$$= 715.5 - 658.3$$

$$= 57.2$$
 - df of SS_{Error}

$$= (n - 1) \times (k - 1)$$

$$= 5 \times 2 = 10$$
 (n은 time 그룹내 샘플 개수, k는 time의 개수)

| ID | 운동전 | 3개월 후 | 6개월 후 | 개인별 평균 |
|--------|------|-------|-------|---------------|
| 1 | 45 | 50 | 55 | 50.0 |
| 2 | 42 | 42 | 45 | 43.0 |
| 3 | 36 | 41 | 43 | 40.0 |
| 4 | 39 | 35 | 40 | 38.0 |
| 5 | 51 | 55 | 59 | 55.0 |
| 6 | 44 | 49 | 56 | 49.7 |
| 기간별 평균 | 42.8 | 45.3 | 49.7 | 전체 평균 45.9 |

반복 측정 분산분석 (Repeated Measure ANOVA)

- One-way ANOVA

| Source | SS | df | MS | F-value | p-value |
|---------|--------|----|-------|---------|---------|
| Between | 143.44 | 2 | 71.72 | 1.504 | 0.254 |
| Within | 715.5 | 15 | 47.7 | | |
| Total | 858.83 | 17 | | | |

- 이렇게 one-way ANOVA를 사용하면 유의하지 않음
- 결론적으로 시간(time)에 따른 차이가 없는 것으로 나타남
- 그러나 이것은 within variance가 비이상적으로 커져서 나타난 현상임

반복 측정 분산분석 (Repeated Measure ANOVA)

- RM ANOVA

| Source | SS | df | MS | F-value | p-value |
|---------------|--------|----|--------|---------|---------|
| Time | 143.44 | 2 | 71.72 | 12.538 | 0.002 |
| Within-groups | 715.5 | 15 | 47.7 | | |
| Subject | 658.3 | 5 | 131.66 | 23.017 | 0.0002 |
| Error | 57.2 | 10 | 5.72 | | |
| Total | 858.83 | 17 | | | |

- One-way ANOVA의 경우와 달리 subject의 within correlation 때문에 비 이상적으로 증가한 within variance를 subject 와 error로 구분하니 시간(time) 그룹과 개인(subject)차이가 유의하게 존재함이 나타남

반복 측정 분산분석 (Repeated Measure ANOVA) ↳ '동분산'과 비슷한 개념.

- 구형성(sphericity) 가정이란
 - RM ANOVA (반복측정 분산분석)에 등장하는 특별한 가정(전제조건)
 - 앞서 이야기 했듯 각 subject가 몇 차례 반복 측정되어 ANOVA의 전제조건인 독립성 가정이 이미 깨졌음
 - 독립성 가정 대신 반복측정 분산분석에서는 구형성(sphericity) 가정이 대신 등장
 - 구형성(Sphericity) 가정: 반복 측정된 자료들의 시차(time)에 따른 분산이 동일
 - Mouchly의 단위행렬 검정을 통해 확인 (P값이 0.05 보다 커야 함)
 - Greenhouse-Geisser ϵ 과 Huynh-Feidt ϵ 은 1에 가까울수록 구형성이 타당함
 - 구형성 가정을 만족하게 되면 일반적인 결과 표를 확인하고, 만족하지 못하면 Greenhouse-Geisser 등의 방법으로 수정된 결과를 이용

반복 측정 분산분석 (Repeated Measure ANOVA)

- 구형성(sphericity) 가정이란
 - 결론적으로 데이터에 포함된 개별 identity들이 어느 정도의 동질성을 확보했는지 확인하는 방법임
 - 만약 매우 병약한 노인과 매우 건강한 청년을 섞어 놓고 신약의 효과를 검정할 경우 개인별 동질성이 매우 낮아 신약의 효과가 개인간 차이에 의해 잘못 검정될 수 있음
 - 이미 독립성 가정이 깨졌으므로, 최대한 독립성과 무작위성을 확보하기 위한 조건으로 생각하면 됨

