

Two-way ANOVA

(이원분산분석)

숙명여자대학교 경영학부 오중산

이원분산분석 소개

[?] 이원분산분석 정의

↳ one way ANOVA도 동일

[?] 두 개 이상 집단간의 종속변수 모평균 차이가 또 다른 독립변수에 의해 영향을 받는지 확인하는

통계분석방법

[?] 두 개 독립변수 간에 상호작용효과(interaction effect)가 존재하는지 확인하는 통계분석방법

[?] 이원(two-way)은 단순히 독립변수가 두 개라는 것이 아니라, 이들이 상호작용한다는 의미

이원분산분석 소개

$IV_1 \rightarrow DV$: 주효과 =

$IV_1 / DV \rightarrow \text{one-way ANOVA}$ Ha 채택

IV_2 : 집단 간 차이에 영향

주효과에 영향

↳ IV_1 과 IV_2 간의 상호작용

[?] 독립변수 간의 상호작용

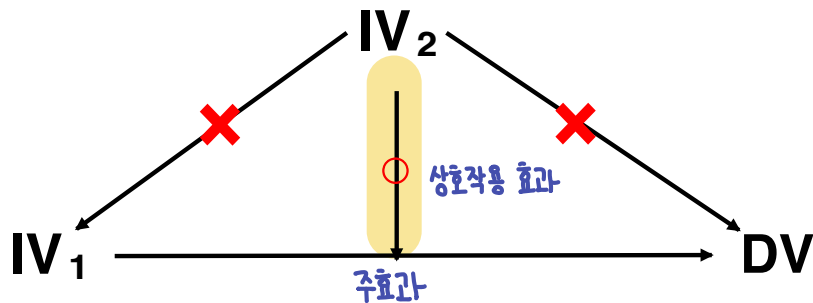
= 집단별 종속변수의 모평균 차이 있음 => 독립변수가 종속변수에 영향 (= 주효과)

[?] 어떤 독립변수(IV_1)가 종속변수(DV)에 미치는 주효과(관계)에 또 다른 독립변수(IV_2)가 영향을 미칠 때 IV_1 와 IV_2 가 상호작용한다고 함

[?] 상호작용은 IV_1 와 DV 간의 기존 ‘관계’에 IV_2 가 영향을 미치는 것을 의미함

[?] IV_1 에 따라 구분된 집단 간에 모평균 차이가 존재할 때, 이 차이가 IV_2 에 의해 영향을 받음

[?] 집단 간의 모평균 차이가 더 벌어지거나, 좁혀질 수 있음



이원분산분석 소개

[?] 이원분산분석 예시

$t=2$

[?] 직무(IV₁: 내근/외근)에 따른 하루 섭취 칼로리 (DV) 모평균 차이에 경력(IV₂: Low/Medium/High)

이 미치는 영향

→ IV₁ - DV : H_a 채택

$$M_{L1} - M_{H1} \neq 0$$

[?] 일원분산분석 결과, 직무에 따른 섭취 칼로리 모평균 차이가 있어야 함



[?] 사전에 One-way ANOVA에서 H_a 채택이 Two-way ANOVA의 전제 조건



: H₀ 채택 → two ANOVA X

[?] 예: 외근직의 섭취 칼로리 모평균이 내근직의 섭취 칼로리 모평균에 비해 큼

[?] 경력에 따른 섭취 칼로리 모평균 차이 여부는 참고사항일뿐 , Two-way ANOVA 전제 조건이 아님

IV₂ → DV : 주효과



$$H_0: M_L = M_M = M_H$$

H_a: 적어도 하나 다름

⇒ 상관없음

이원분산분석 소개

[?] 이원분산분석 예시 one-way ANOVA를 통해 확인된 전제조건
: $M_H > M_M$

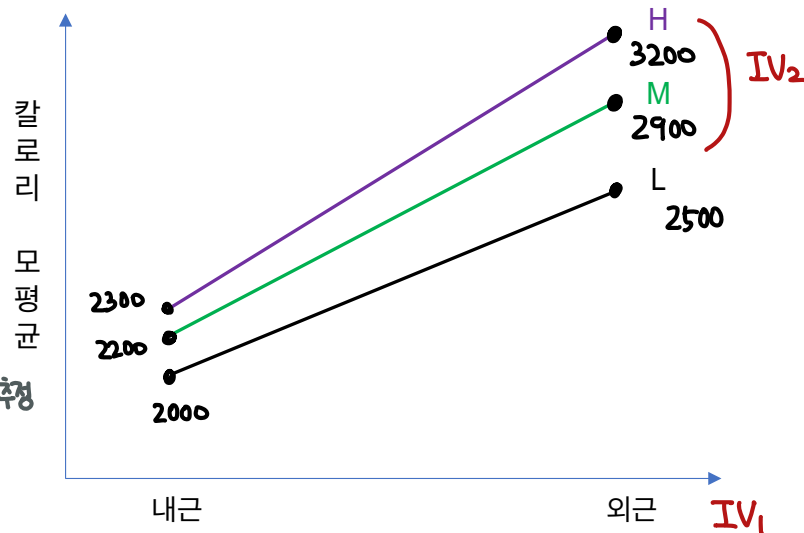
[?] 직무(IV_1)에 따라 하루 섭취 칼로리(DV) 모평균 차이에 경력(IV_2)이 미치는 영향

[?] IV_2 가 기존 관계를 강화하는 경우

[?] 경력이 높아질수록 외근직과 내근직의 하루 섭취 칼로리 모평균 차이가 더 확대됨

	경력L	경력M	경력H	평균
내근직	2,000	2,200	2,300	2,167
외근직	2,500	2,900	3,200	2,867
차이	500	700	900	700

표본평균 가지고 모평균 추정



이원분산분석 소개

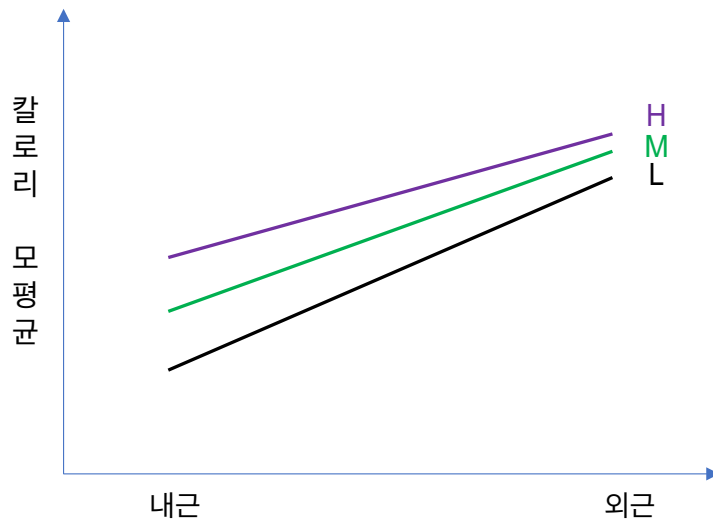
[?] 이원분산분석 예시

[?] 직무(IV₁)에 따라 하루 섭취 칼로리(DV) 모평균 차이에 경력(IV₂)이 미치는 영향

[?] IV₂가 기존 관계를 약화하는 경우 : 내근직에 비해서 외근직의 종속변수 모평균이 더 큰데, 그 차이는 경력이 커짐에 따라 점점 줄어든다

[?] 경력이 높아질수록 외근직과 내근직의 하루 섭취 칼로리 모평균 차이는 줄어듦

	경력L	경력M	경력H	평균
내근직	2,000	2,200	2,300	2,167
외근직	2,500	2,600	2,650	2,583
차이	500	400	350	417



이원분산분석 가설과 세 가지 전제조건

[?] 이원분산분석의 두 가지 가설

[?] H_0 : 두 독립변수 간에 상호작용효과가 없다.

[?] H_a : 두 독립변수 간에 상호작용효과가 있다.

↳ 강화할 수도, 약화할 수도

[?] 세 가지 전제조건

[?] 독립성은 기본적으로 만족해야 하며, 정규성/등분산성은 IV_1 을 기준으로 구분된 집단에 대해서만 확인

이원분산분석 가설검정

① F_A 유의 : 전제조건

② F_{AC} : 상호작용 효과 있는지 없는지

[?] Two-way ANOVA Table과 가설검정

[?] F_A 에 따른 p -value와 F_{AC} 에 따른 p -value가 모두 유의하면 대립가설 채택

[?] 두 p -value 중에서 하나라도 유의하지 않으면 귀무가설 채택

[?] F_A 와 F_C 는 두 독립변수 각각에 대한 One-way ANOVA 가설검정에 활용

[?] $SSTR(\text{요인효과}) = SSTR_A + SSTR_C + SSTR_{AC}$

분산요인	제곱 합	자유도	평균자승	F-statistics
A(IV₁) : A의 주효과	$SSTR_A$	(k-1)	$MSTR_A = SSTR_A / (k-1)$	$F_A = MSTR_A / MSE$
C(IV₂) : C의 주효과	$SSTR_C$	(g-1)	$MSTR_C = SSTR_C / (g-1)$	$F_C = MSTR_C / MSE$
A×C(IV₁ × IV₂)	$SSTR_{AC}$	(k-1)(g-1)	$MSTR_{AC} = SSTR_{AC} / (k-1)(g-1)$	$F_{AC} = MSTR_{AC} / MSE$
오차분산 <small>↗ 상호작용 효과</small>	SSE	<small>전제검정</small> $N-k \times g$	$MSE = SSE / [N-k \times g]$	
총분산	SST	N - 1		

이원분산분석 검정 절차

IV₁ : gender

IV₂ : OS

DV: expense

[?] 일원분산분석의 1~5단계 수행 one-way ANOVA

[?] IV₁에 따라 구분된 집단 간에 DV 모평균 차이가 유의함을 확인해야 함

[?] 6단계는 목적에 따라 수행할 수도 있고, 안 할 수도 있음 : 목적 [IV₁-DV : one-way ⇒ 함
↳ 사후검정

[?] IV₁에 따른 일원분산분석도 목적이라면 수행

상호작용 : two-way ⇒ 안함

[?] 7단계: 가설검정 및 그래프 그리기

[?] aov 함수를 이용한 이원분산분석 가설검정

[?] 기본 명령문: $\text{aov}(\text{DV} \sim \text{IV}_1 * \text{IV}_2, \text{data} = \text{이상치 제거된 전체 df})$

[?] HH패키지에 있는 interaction2wt 함수 사용하여 그래프 그리기

[?] 기본 명령문: $\text{interaction2wt}(\text{DV} \sim \text{IV}_1 * \text{IV}_2, \text{data} = \text{이상치 제거된 전체 df})$



이원분산분석 검정 절차

[?] 8단계: 추가 분석

내·외	L·M·H	내·L
IV ₁	IV ₂	내·M
		⋮
		6개

⇒ one-way ANOVA

[?] IV₁과 IV₂를 동시에 고려하여 집단을 세분화했을 때, 집단 간에 DV 모평균의 차이가 존재할까?

[?] 이원분산분석에서 대립가설이 채택되면 이런 추가적인 일원분산분석을 추가 수행할 수 있음

[?] 주의사항! 이원분산분석에서 대립가설의 채택여부와 무관하게 추가 일원분산분석에서 집단 간에 종속변수 모평균 차이가 유의하게 추정될 수 있음 ⇒ 귀무가설 채택되어도 추가 분석에서는 대립가설 채택될 수 있음

[?] IV₁과 IV₂를 동시에 고려한 새로운 변수 (IV₃)를 만들고 이에 따라 일원분산분석 1~6단계 시행

[?] IV₁에 따른 집단 개수 k와 IV₂에 따른 집단 개수 g를 고려하면 IV₃에 따른 집단 개수는 k×g가 됨

이원분산분석 실습

[?] 다음과 같은 Two-way ANOVA를 실행하시오.

[?] 데이터: pttest

[?] IV_1 : gender / IV_2 : os

[?] DV: expense

[?] 유의수준(α) = 0.05

STEP 1. 가설수립

사전작업

- ① `table(is.na(two_anova $ expense))` : 종속변수에 na 있는지 확인
- ② `two_anova %>% group_by(gender) %>% summarise(mean(expense, na.rm=T))`
- H₀: 두 독립변수 간에 상호작용 효과가 없다 H_a: 두 독립변수 간에 상호작용 효과가 있다.

이상치 검토 및 제거

- ① `library(psych)` ② `descr <- describe(two_anova $ expense)`
- ③ `descr <- descr %>% mutate(UL = mean + 3 * sd, LL = mean - 3 * sd)` ④ `table(two_anova $ expense > descr $ UL)`
- ⑤ `two_anova_new <- two_anova %>% filter(expense <= descr $ UL)`

STEP2: 서브 데이터프레임 만들기

- ① `two_anova_male <- two_anova_new %>% filter(gender == "Male")`
- ② `two_anova_female <- two_anova_new %>% filter(gender == "female")`

STEP3: 정규성 검토

- ① `summary(two_anova_male)` ② `hist(two_anova_male $ expense, breaks = seq(0, 2000, 40))`
- ③ `Shapiro.test(two_anova_male $ expense)`

STEP 4. 등분산성 검토

- ① `library(car)` ② `leveneTest(expense ~ gender, data = two_anova_new)`

```
> leveneTest(expense ~ gender, data = two_anova_new)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  1  10.491 0.001238 **
---
1035
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d = 0.05 보다 작음 => H_a 채택 => 등분산성 만족 X

STEP5 : 이분산 가정 one-way ANOVA 시험 (welch test 시험)

`oneway.test(expense ~ gender, data = two_anova_new)`

```
> oneway.test(expense~gender, data = two_anova_new)

One-way analysis of means (not assuming equal variances)

data: expense and gender
F = 51.752, num df = 1.0, denom df = 1004.8, p-value = 1.23e-12
```

```
> two_anova_new %>% group_by(gender) %>% summarise(mean(expense))
# A tibble: 3 x 2
  gender mean(expense)
  <chr>      <dbl>
1 Female    553.
2 Male     359.
3 NA       303.
```

=> m(female) > m(male)

기준으로 집단 구분했을 때, 두 집단간의 종속변수 모평균의 차이 존재

STEP 6 사후분석 생략

=> 뮤(female) > 뮤(male)

- ① two-way ANOVA가 목적
- ② 집단 2개 (3개 이상 사후분석)

STEP 7: two-way ANOVA 시행 및 그래프 그리기

- ① `two_anova_result <- aov(expense ~ gender * OS, data = two_anova_new)`
- ② `summary(two_anova_result)`

```
> summary(two_anova_result)
          Df Sum Sq Mean Sq F value
gender IV1      1  9556293  9556293   51.48
os       IV2      1  2041400  2041400   11.00
gender:os IV, X IV, 1 2367490 2367490   12.75
Residuals 1033 191746440 185621
          Pr(>F)
gender 1.38e-12 ***
os      0.000944 ***
gender:os 0.000372 ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
결측으로 인하여 7개의 관측치가 삭제되었습니다.
```

- ③ `library(HH)`
- ④ `table(two_anova_new$gender)`
- ⑤ `two_anova_new %>% group_by(gender) %>% summarise(mean(expense))`
- ⑥ `two_anova_new$gender <- factor(two_anova_new$gender, levels = c("Male", "Female"))`
- ⑦ `interaction2wt(expense ~ gender * OS, data = two_anova_new)`

IV2 확인안해도 됨 (one way ANOVA)

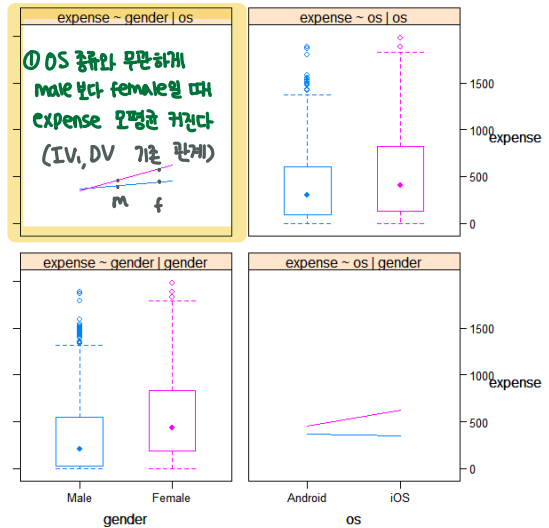
만약 하라하면

`two_anova_new %>% group_by(OS) %>%`

`summarise(mean(expense))`

안드로이드면 OS가 iOS면
차이가 모평균 차이 ← male, female 모평균 차이가 더 커짐
별로 크지 않음
OS 및 쓰냐에 따라 차이 더 벌어질 수도
⇒ 기존의 관계 강화
☆ gender에 따른 구분했을 때 남성보다 여성의 expense 모평균 더 크다 라는 기존의 관계에 OS가 영향 미침 ; OS: 성별에 따른 차이 더 커짐 안드로이드 : 완화

expense: main effects and 2-way interactions



+

STEP 8: 집단을 네 개로 세분화 했을 때

집단 네 개로 구분하는 변수 만들기

```
## 집단 네 개: MA, Mi, FA, Fi ##
two_anova_new <- two_anova_new %>% mutate(genderos = ifelse(gender == "Male" & os == "Android", "MA", ifelse(gender == "Male" & os == "iOS", "Mi", ifelse(gender == "Female" & os == "Android", "FA", "Fi"))))
```

서브데이터 프레임 만들기

```
two_anova_MA <- two_anova_new %>% filter(genderos == "MA")
two_anova_Mi <- two_anova_new %>% filter(genderos == "Mi")
two_anova_FA <- two_anova_new %>% filter(genderos == "FA")
two_anova_Fi <- two_anova_new %>% filter(genderos == "Fi")
```

정규성 검토와 등분산성 검토

`shapiro.test(two_anova_MA$expense)`

`leveneTest(expense ~ genderos, data = two_anova_new)`

p-value: 2.376e-05 < α ⇒ 등분산성 만족 X

이분산 가정 One-way ANOVA 시행

oneway.test(expense ~ genders, data = two_anova_new)

종속 독립

```
> oneway.test(expense~genders, data = two_anova_new)

One-way analysis of means (not assuming equal variances)

data: expense and genders
F = 22.5, num df = 3.00, denom df = 545.95, p-value = 9.516e-14
```

> d ⇒ H_a 채택 ⇒ 적어도 하나의 집단의 expense 평균은 다른 집단과 다른데, 어떻게 다른지 사후분석 해야함

사후분석 시행

* 등분산 가정 ⇒ dunn.test 진행

① library(dunn.test)

이분산 가정 ⇒ dunn.test

@dunn.test(two_anova_new \$ expense, two_anova_new \$ genders, method = "bonferroni")

DV (종속변수)

IV (독립변수)

Col	Mean-			
Row	Mean		FA	Fi
Fi	-3.722712			
			0.0006*	<0.025 (알)
MA	3.374870			
			0.0022*	0.0000*
Mi	3.533885			
			0.0012*	0.0000*
				0.403644
				1.0000

alpha = 0.05
Reject H₀ if p <= alpha/2

μ(F_A) < μ(F_i) : p-value가 0.0006 < 0.025 (알파/2)고
t 값이 -3.723 (마이너스)로 나왔으므로

μ(F_A) > μ(MA), μ(F_A) > μ(Mi), μ(F_i) > μ(MA)
μ(F_i) > μ(Mi), μ(MA) = μ(Mi)

⇒ F_i > F_A > MA = M_i

↳ t 검정은 양측검정

각각 쌍을 지어서 t 검정