Repeated Measures ANOVA-test

Youngjin\_Lee

2022 2 13

Table of Contents

# 반복측정 분산분석(Repeated Measures ANOVA)

통계분석 차이검정 : 반복측정 분산분석(Repeated Measures ANOVA) test 문제 K대학에서 운영하는 6개월 어학 프로그램이 있다. 어학 프로그램의 효과를 측정하기 위해 학습 프로그램에 참여한 학생을 대상으로 프로그램 참여전, 3개월 후, 6개월 후에 영어실력을 테스트하였다. 과련 어학프로그램은 효과가 있는지? 있다면 언제부터 효과가 나타났을지를 검증해 보자

# 1.기본 package 설정, library 로드

# 2.데이터 불러오기

rma\_tb <- read\_csv('data\\RMA.csv',   
 col\_names = TRUE,  
 locale=locale('ko', encoding='euc-kr'), # 한글  
 na=".") %>%  
 mutate\_if(is.character, as.factor) %>%  
 mutate(시점 = factor(시점,  
 levels=c(1:3),  
 labels=c("사전","3개월","6개월")))

## Rows: 135 Columns: 3  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## dbl (3): id, 시점, 점수  
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(rma\_tb)

## spec\_tbl\_df [135 x 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ id : num [1:135] 1 1 1 2 2 2 3 3 3 4 ...  
## $ 시점: Factor w/ 3 levels "사전","3개월",..: 1 2 3 1 2 3 1 2 3 1 ...  
## $ 점수: num [1:135] 63 64 68 60 61 67 61 67 70 57 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. id = col\_double(),  
## .. 시점 = col\_double(),  
## .. 점수 = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

rma\_tb

## # A tibble: 135 x 3  
## id 시점 점수  
## <dbl> <fct> <dbl>  
## 1 1 사전 63  
## 2 1 3개월 64  
## 3 1 6개월 68  
## 4 2 사전 60  
## 5 2 3개월 61  
## 6 2 6개월 67  
## 7 3 사전 61  
## 8 3 3개월 67  
## 9 3 6개월 70  
## 10 4 사전 57  
## # ... with 125 more rows

# 3.기술통계분석

skim(rma\_tb)

Data summary

|  |  |
| --- | --- |
| Name | rma\_tb |
| Number of rows | 135 |
| Number of columns | 3 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 1 |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| 시점 | 0 | 1 | FALSE | 3 | 사전: 45, 3개월: 45, 6개월: 45 |

**Variable type: numeric**

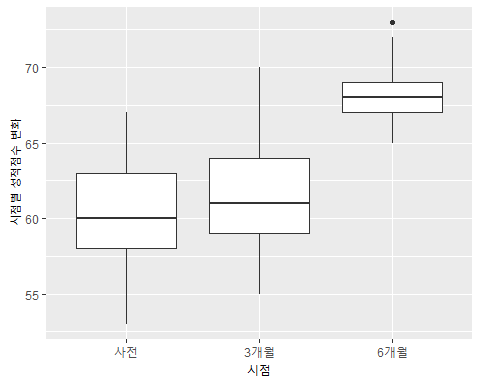
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | 0 | 1 | 23.00 | 13.04 | 1 | 12 | 23 | 34 | 45 | ▇▇▇▇▇ |
| 점수 | 0 | 1 | 63.31 | 4.56 | 53 | 59 | 64 | 67 | 73 | ▃▇▅▇▂ |

rma\_tb %>%   
 group\_by(시점) %>%  
 get\_summary\_stats(점수)

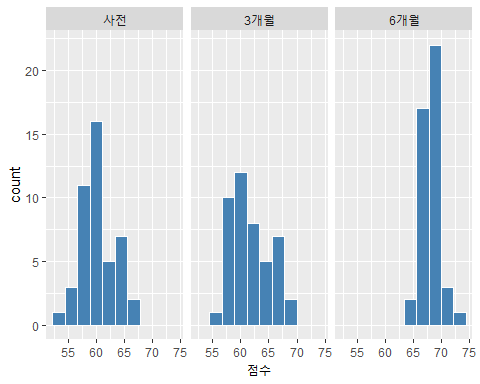
## # A tibble: 3 x 14  
## 시점 variable n min max median q1 q3 iqr mad mean sd  
## <fct> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 사전 점수 45 53 67 60 58 63 5 2.96 60.2 3.22  
## 2 3개월 점수 45 55 70 61 59 64 5 4.45 61.6 3.60  
## 3 6개월 점수 45 65 73 68 67 69 2 1.48 68.1 1.83  
## # ... with 2 more variables: se <dbl>, ci <dbl>

# 4.그래프 그리기(박스그래프,히스토그램)

rma\_tb %>%   
 ggplot(mapping = aes(x = 시점,  
 y = 점수)) +  
 geom\_boxplot() +  
 labs(y = "시점별 성적점수 변화")



rma\_tb %>%   
 ggplot(mapping = aes(x = 점수)) +  
 geom\_histogram(bins = 10, # binwidth=1,000: 값차이  
 color = "white",   
 fill = "steelblue") +  
 facet\_wrap(~ 시점) # 그룹별 분리



# 5.이상치 제거, 이상치 확인

rma\_tb %>%  
 group\_by(시점) %>%  
 identify\_outliers(점수)

## # A tibble: 1 x 5  
## 시점 id 점수 is.outlier is.extreme  
## <fct> <dbl> <dbl> <lgl> <lgl>   
## 1 6개월 37 73 TRUE FALSE

# 6.정규분포 검정

rma\_tb %>%  
 group\_by(시점) %>%  
 shapiro\_test(점수)

## # A tibble: 3 x 4  
## 시점 variable statistic p  
## <fct> <chr> <dbl> <dbl>  
## 1 사전 점수 0.972 0.339   
## 2 3개월 점수 0.962 0.149   
## 3 6개월 점수 0.954 0.0737

# 7.구형성(sphericity)검정: Mauchly’s test.

rma\_result <- rma\_tb %>%  
 rstatix::anova\_test(dv = 점수,   
 wid = id,   
 within = 시점)  
rma\_result

## ANOVA Table (type III tests)  
##   
## $ANOVA  
## Effect DFn DFd F p p<.05 ges  
## 1 시점 2 88 248.614 6.24e-37 \* 0.579  
##   
## $`Mauchly's Test for Sphericity`  
## Effect W p p<.05  
## 1 시점 0.963 0.446   
##   
## $`Sphericity Corrections`  
## Effect GGe DF[GG] p[GG] p[GG]<.05 HFe DF[HF] p[HF]  
## 1 시점 0.964 1.93, 84.88 1.04e-35 \* 1.008 2.02, 88.69 6.24e-37  
## p[HF]<.05  
## 1 \*

# 8.사후검정(Multicamparison test )

## Fisher LSD

rma\_tb %>%   
 pairwise\_t\_test(점수 ~ 시점,  
 paired=TRUE,  
 p.adj="bonferroni")

## # A tibble: 3 x 10  
## .y. group1 group2 n1 n2 statistic df p p.adj p.adj.signif  
## \* <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 점수 사전 3개월 45 45 -3.44 44 1 e- 3 4 e- 3 \*\*   
## 2 점수 사전 6개월 45 45 -23.0 44 3.48e-26 1.04e-25 \*\*\*\*   
## 3 점수 3개월 6개월 45 45 -17.0 44 6.18e-21 1.85e-20 \*\*\*\*

# 부록: 일변량 ANOVA 검정

rma\_result\_ow <- aov(점수 ~ 시점+Error(id/시점),   
 data=rma\_tb)  
summary(rma\_result\_ow)

##   
## Error: id  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Residuals 1 13.24 13.24   
##   
## Error: id:시점  
## Df Sum Sq Mean Sq  
## 시점 2 1136 568  
##   
## Error: Within  
## Df Sum Sq Mean Sq F value Pr(>F)   
## 시점 2 500.4 250.2 28.43 5.9e-11 \*\*\*  
## Residuals 129 1135.4 8.8   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# 부록: car 이용 구형성(sphericity)검정: Mauchly’s test.

library(car)

## 필요한 패키지를 로딩중입니다: carData

##   
## 다음의 패키지를 부착합니다: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

rma.matrix <- cbind(rma\_tb$점수[rma\_tb$시점=="사전"],   
 rma\_tb$점수[rma\_tb$시점=="3개월"],   
 rma\_tb$점수[rma\_tb$시점=="6개월"])  
head(rma.matrix)

## [,1] [,2] [,3]  
## [1,] 63 64 68  
## [2,] 60 61 67  
## [3,] 61 67 70  
## [4,] 57 61 68  
## [5,] 58 61 68  
## [6,] 58 62 68

rma.model.lm <- lm(rma.matrix ~ 1) # ~1 : 데이터를 하나로 묶기  
time.f <- factor(c("사전","3개월","6개월"))   
options(contrasts=c("contr.sum", "contr.poly"))  
rma.result.mt <- Anova(rma.model.lm, #Anova 대문자  
 idata=data.frame(time.f),  
 idesign=~time.f,   
 type="III")  
  
summary(rma.result.mt, multivariate=F)

## Warning in summary.Anova.mlm(rma.result.mt, multivariate = F): HF eps > 1  
## treated as 1

##   
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity  
##   
## Sum Sq num Df Error SS den Df F value Pr(>F)   
## (Intercept) 541120 1 885.6 44 26884.92 < 2.2e-16 \*\*\*  
## time.f 1614 2 285.6 88 248.61 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
##   
## Mauchly Tests for Sphericity  
##   
## Test statistic p-value  
## time.f 0.96318 0.44641  
##   
##   
## Greenhouse-Geisser and Huynh-Feldt Corrections  
## for Departure from Sphericity  
##   
## GG eps Pr(>F[GG])   
## time.f 0.96449 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## HF eps Pr(>F[HF])  
## time.f 1.007868 6.236035e-37