

다변량

4장 과제

과목	다변량데이터분석
담당교수	임태진 교수님
전공	산업정보시스템공학과
학번	20201368
이름	한채원
제출일	2022.10.26

car 패키지의 Baumann 데이터를 사용하여 다음을 수행하시오.

단, group 열의 Basal=전통적인 교수법, DRTA = 혁신적인 방법, Strat = 또 다른 혁신적인 방법을 나타낸다.

1. 5개 시험 성적 분포를 비교하는 상자그림을 하나의 그래프로 작성하고 분석하시오.

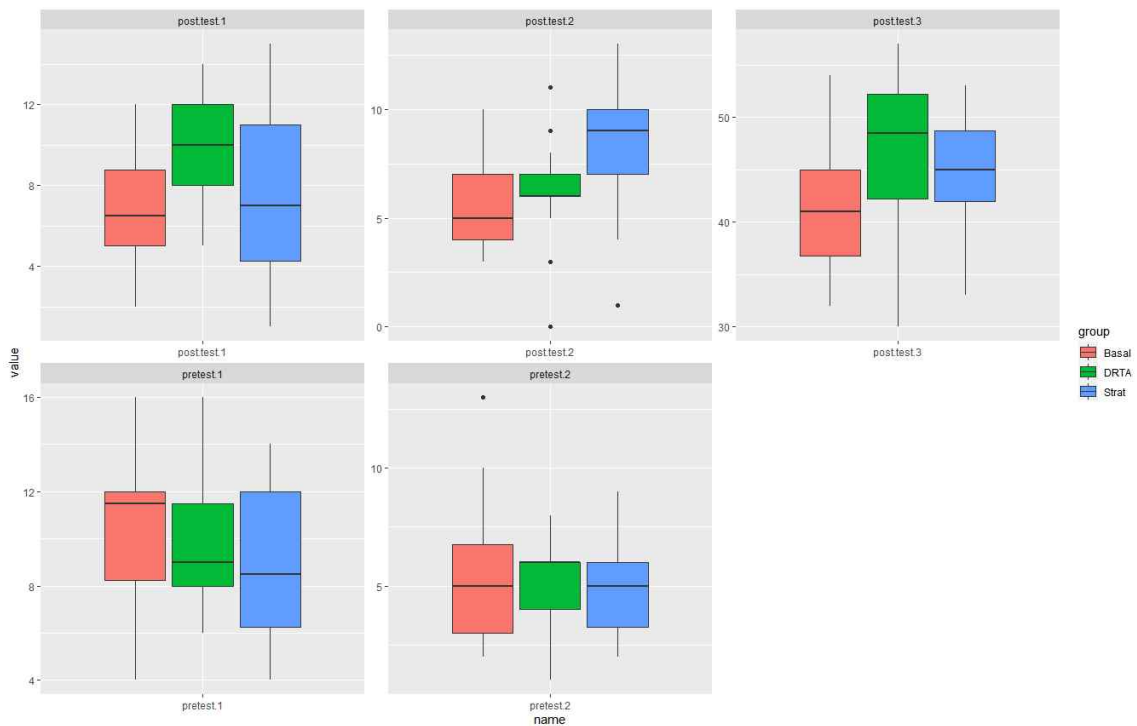
우선 car패키지에서 Baumann을 불러온다.

```
> library(car)
> str(Baumann)
'data.frame':   66 obs. of  6 variables:
 $ group      : Factor w/ 3 levels "Basal","DRTA",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ pretest.1  : int  4 6 9 12 16 15 14 12 12 8 ...
 $ pretest.2  : int  3 5 4 6 5 13 8 7 3 8 ...
 $ post.test.1: int  5 9 5 8 10 9 12 5 8 7 ...
 $ post.test.2: int  4 5 3 5 9 8 5 5 7 7 ...
 $ post.test.3: int  41 41 43 46 46 45 45 32 33 39 ...
```

box plot을 다음과 같이 작성한다. 스케일링을 해준다.

```
> library(tidyverse)
> ldf=pivot_longer(Baumann,2:6)
> str(ldf)
tibble [330 x 3] (S3: tbl_df/tbl/data.frame)
 $ group: Factor w/ 3 levels "Basal","DRTA",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ name : chr [1:330] "pretest.1" "pretest.2" "post.test.1" "post.test.2" ...
 $ value: int [1:330] 4 3 5 4 41 6 5 9 5 41 ...

> g = ggplot(ldf,aes(x=name, y=value))+geom_boxplot(aes(fill=group))
> g+facet_wrap(vars(name),scales='free')
> bau = Baumann[,-1]
```



다음과 같은 결과가 나오는 것을 알 수 있다. 교수법 적용 이전 Basal 교수법이 평균과 분포 가장 큼을 알 수 있다. 하지만 교수법 이후 DRTA 교수법이 pretest에서보다 눈에 띄는 결과 값을 가짐을 알 수 있다. 평균과 최댓값, 최솟값 값들이 비교적 높다는 것을 파악할 수 있다. 특히 post3에서 value값이 다른 테스트보다 유독 큰 것을 알 수 있다.

2. 5개 시험 성적의 공분산행렬이 교수법에 따라 차이가 유의미한지 검정하시오.

```
> library(biotools)
> bau.bm = boxM(Baumann[, -1], Baumann[[1]])
> bau.bm
Box's M-test for Homogeneity of Covariance Matrices
```

```
data: Baumann[, -1]
Chi-Sq (approx.) = 29.557, df = 30, p-value = 0.4885
```

코드 결과 p값이 0.485로 이는 유의확률 0.05보다 크므로 공분산 행렬이 동일하다는 가설을 기각할 수 없으므로 다변량 접근 방식의 동질검정을 만족하게 된다. 따라서 교수법에 따라 공분산 행렬의 차이가 유의미하다고 볼 수 없다.

3. 교수법 적용 이전 (pretest) 2개 시험 성적 평균벡터가 교수법에 따라 차이가 유의한지 검정하시오.

```
> bau.pre = manova(as.matrix(bau[1:2])~group, data=Baumann)
```

```
> summary(bau.pre)
```

```
      Df  Pillai approx F num Df den Df Pr(>F)
group    2 0.034706  0.55628    4   126 0.6948
Residuals 63
```

```
> summary(bau.pre, test="Wilks")
```

```
      Df  Wilks approx F num Df den Df Pr(>F)
group    2 0.96529  0.55237    4   124 0.6976
Residuals 63
```

```
> summary(bau.pre, test="Hotell")
```

```
      Df Hotelling-Lawley approx F num Df den Df Pr(>F)
group    2      0.035954  0.5483    4   122 0.7006
Residuals 63
```

```
> summary(bau.pre, test="Roy")
```

```
      Df      Roy approx F num Df den Df Pr(>F)
group    2 0.035954  1.1325    2    63 0.3287
Residuals 63
```

다음과 같이 평균벡터가 교수법에 따라 차이가 있는지 알아보기 위해 Pillai, Wilks, Hotelling, Roy 검정을 실시했다. 위 4개의 검정에서 p 값이 모두 유의확률 0.05보다 크므로 평균간의 유의미한 차이가 없다는 가설을 기각할 수 없다. 따라서 교수법 적용 이전 (pretest) 2개 시험 성적 평균벡터가 교수법에 따라 차이가 유의미하다고 볼 수 없다.

4. 교수법 적용 이후(posttest) 3개 시험 성적 평균벡터가 교수법에 따라 차이가 유의한지 검토하시오.

```
> bau.pos = manova(as.matrix(bau[3:5])~group,data=Baumann)
```

```
> summary(bau.pos)
```

```
      Df  Pillai approx F num Df den Df Pr(>F)
group    2 0.40825  5.3005    6   124 6.765e-05 ***
Residuals 63
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(bau.pos, test="Wilks")
```

```
      Df  Wilks approx F num Df den Df Pr(>F)
group    2 0.63202  5.2433    6   122 7.774e-05 ***
Residuals 63
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(bau.pos, test="Hotell")
```

```
      Df Hotelling-Lawley approx F num Df den Df Pr(>F)
group    2      0.51852  5.1852    6   120 8.949e-05 **
Residuals 63
```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(bau.pos, test="Roy")
      Df      Roy approx F num Df den Df      Pr(>F)
group   2 0.31845   6.5813     3   62 0.0006206 ***
Residuals 63
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

다음과 같이 평균벡터가 교수법에 따라 차이가 있는지 알아보기 위해 Pillai, Willks, Hotelling, Roy 검정을 실시했다. 위 4개의 검정에서 p 값이 모두 유의확률 0.05보다 작으므로 평균간의 유의미한 차이가 없다는 가설을 기각할 수 있다. 따라서 교수법 적용 이전 (posttest) 3개 시험 성적 평균벡터가 교수법에 따라 차이가 유의미하다고 볼 수 있다.