1-1) 분산분석표 구하기

```
data1 <- scan(what=list(1,1))
1 8.44 1 8.36 1 8.28
2 8.59 2 8.91 2 8.60
3 9.34 3 9.41 3 9.69
4 8.92 4 8.92 4 8.74
df1 <- data.frame(data1)
df1$trt <- as.factor(df1$trt)
result <- lm(y~trt,data=df1)
```

```
> anova(result)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
trt        3 1.9788 0.65960  31.187 9.171e-05 ***
Residuals  8 0.1692 0.02115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1-2) 기각역 직접 계산

```
> qf(0.05, 3, 8, lower.tail=FALSE)
[1] 4.066181
```

31.187>$F_{0.05,3,8} \simeq 4.07$이므로 귀무가설 기각

1-3) 다중비교

PDIFF 함수를 이용하기 위해 아래 패키지 불러오기

library(sasLM)

(1) 최소유의차

```
> PDIFF(y~trt, df1, adj="lsd")
      Estimate Lower CL Upper CL Std. Error t value Df  Pr(>|t|)
1 - 2    -0.34 -0.61382 -0.06618    0.11874 -2.8633  8 0.0210439 *
1 - 3    -1.12 -1.39382 -0.84618    0.11874 -9.4321  8 1.311e-05 ***
1 - 4    -0.50 -0.77382 -0.22618    0.11874 -4.2108  8 0.0029528 **
2 - 3    -0.78 -1.05382 -0.50618    0.11874 -6.5688  8 0.0001750 ***
2 - 4    -0.16 -0.43382  0.11382    0.11874 -1.3474  8 0.2147532
3 - 4     0.62  0.34618  0.89382    0.11874  5.2213  8 0.0008013 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

유의수준 5% 하에서 1-2, 1-3, 1-4, 2-3, 3-4 간 차이가 유의함

평균 차이의 절댓값과 LSD 기준값을 비교하기 위해 아래와 같이 직접 구할 수 있음

```
> qt(0.975, 8)*sqrt(0.02115)*sqrt(1/3+1/3)
[1] 0.2738228
```

$LSD \simeq 0.27$로 Estimate의 절댓값이 이보다 큰 1-2, 1-3, 1-4, 2-3, 3-4 간 차이가 유의하다는 것을 다시 확인할 수 있음

(2) Bonferroni

```
> PDIFF(y~trt, df1, adj="bon")
      Estimate Lower CL Upper CL Std. Error t value Df  Pr(>|t|)
1 - 2    -0.34 -0.75309  0.07309    0.11874 -2.8633  8  0.126263
1 - 3    -1.12 -1.53309 -0.70691    0.11874 -9.4321  8 7.866e-05 ***
1 - 4    -0.50 -0.91309 -0.08691    0.11874 -4.2108  8  0.017717 *
2 - 3    -0.78 -1.19309 -0.36691    0.11874 -6.5688  8  0.001050 **
2 - 4    -0.16 -0.57309  0.25309    0.11874 -1.3474  8  1.000000
3 - 4     0.62  0.20691  1.03309    0.11874  5.2213  8  0.004808 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

유의수준 5% 하에서 1-3, 1-4, 2-3, 3-4 간 차이가 유의함
평균 차이의 절댓값과 MSD 기준값을 비교하기 위해 아래와 같이 직접 구할 수 있음

```
> qt(0.025/6, 8, lower.tail = FALSE)*sqrt(0.02115)*sqrt(1/3+1/3)
[1] 0.413094
```

$MSD \simeq 0.41$로 Estimate의 절댓값이 이보다 큰 1-3, 1-4, 2-3, 3-4 간 차이가 유의하다는 것을 다시 확인할 수 있음

(3) Scheffe

```
> PDIFF(y~trt, df1, adj="scheffe")
      Estimate Lower CL Upper CL Std. Error F value Df2      Pr(>F)
1 - 2    -0.34 -0.75473  0.07473    0.11874  2.7329   8 0.1136268
1 - 3    -1.12 -1.53473 -0.70527    0.11874 29.6548   8 0.0001102 ***
1 - 4    -0.50 -0.91473 -0.08527    0.11874  5.9102   8 0.0199216 *
2 - 3    -0.78 -1.19473 -0.36527    0.11874 14.3830   8 0.0013773 **
2 - 4    -0.16 -0.57473  0.25473    0.11874  0.6052   8 0.6298908
3 - 4     0.62  0.20527  1.03473    0.11874  9.0875   8 0.0058968 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

유의수준 5% 하에서 1-3, 1-4, 2-3, 3-4 간 차이가 유의함
평균 차이의 절댓값과 Scheffe 기준값을 비교하기 위해 아래와 같이 직접 구할 수 있음

```
> sqrt(3*qf(0.05, 3, 8, lower.tail=FALSE))*sqrt(0.02115)*sqrt(1/3+1/3
[1] 0.4147281
```

$Scheffe \simeq 0.41$로 Estimate의 절댓값이 이보다 큰 1-3, 1-4, 2-3, 3-4 간 차이가 유의하다는 것을 다시 확인할 수 있음

(4) Tukey

```
> PDIFF(y~trt, df1, adj="tukey")
      Estimate Lower CL Upper CL Std. Error t value Df  Pr(>|t|)
1 - 2    -0.34 -0.72026  0.04026    0.11874 -2.8633  8 0.0806835 .
1 - 3    -1.12 -1.50026 -0.73974    0.11874 -9.4321  8 6.095e-05 ***
1 - 4    -0.50 -0.88026 -0.11974    0.11874 -4.2108  8 0.0125852 *
2 - 3    -0.78 -1.16026 -0.39974    0.11874 -6.5688  8 0.0007942 ***
2 - 4    -0.16 -0.54026  0.22026    0.11874 -1.3474  8 0.5615203
3 - 4     0.62  0.23974  1.00026    0.11874  5.2213  8 0.0035418 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

유의수준 5% 하에서 1-3, 1-4, 2-3, 3-4 간 차이가 유의함
평균 차이의 절댓값과 Tukey 기준값을 비교하기 위해 아래와 같이 직접 구할 수 있음

```
> 4.53/sqrt(2)*sqrt(0.02115)*sqrt(1/3+1/3)
[1] 0.3803582
```

$Tukey \simeq 0.38$로 Estimate의 절댓값이 이보다 큰 1-3, 1-4, 2-3, 3-4 간 차이가 유의하다는 것을 다시 확인할 수 있음

+ TukeyHSD 함수 이용

result <- aov(y~trt,data=df1)

TukeyHSD(result)

```
> TukeyHSD(result)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = y ~ trt, data = df1)

$trt
     diff         lwr       upr     p adj
2-1  0.34 -0.04025823 0.7202582 0.0806835
3-1  1.12  0.73974177 1.5002582 0.0000609
4-1  0.50  0.11974177 0.8802582 0.0125852
3-2  0.78  0.39974177 1.1602582 0.0007942
4-2  0.16 -0.22025823 0.5402582 0.5615203
4-3 -0.62 -1.00025823 -0.2397418 0.0035418
```

TukeyHSD 함수를 이용해도 유의확률이 위의 결과와 동일한 것을 확인할 수 있음

1-4) 잔차 직접 계산

```
> mean(df1[df1$trt==1,]$y)
[1] 8.36
> 8.28-8.36
[1] -0.08
```

위에서 구한 잔차를 e 변수에 저장하고 studentized 잔차 계산
(아래에 계속)

```
> e=-0.08
> e/sqrt(2*0.02115/3)
[1] -0.6737215
```

2. 선형대비

```
data2 <- c(0,2,1,3,1,2,3,4,1,5,
           1,3,4,6,8,7,5,3,4,5,
           14,26,25,18,19,22,21,16,20,30)
trt <- rep(1:3, each=10)
df2 <- data.frame(y=data2, trt=as.factor(trt))
```

PDIFF 함수를 이용하기 위해 아래 패키지 불러오기

library(sasLM)

```
> ESTM(t(c(0,2,-1,-1)), y~trt, df2)
     Estimate  Lower CL  Upper CL Std. Error   t value Df     Pr(>|t|)
[1,]    -21.3 -26.33431 -16.26569   2.453569 -8.681232 27 2.691193e-09
attr(,"Estimability")
[1] TRUE
> (-8.681232)^2
[1] 75.36379
> qf(0.05, 1, 27, lower.tail=FALSE)
[1] 4.210008
```

위의 ESTM 결과의 t value의 제곱은 약 75.364이고 $F_{0.05,1,27}$은 약 4.21

75.364>4.21이므로 유의수준 5% 하에서 귀무가설 기각

3. Hartley

```
> var(df3[df3$trt=="A",]$y)
[1] 1.1
> var(df3[df3$trt=="B",]$y)
[1] 15.76667
> 15.76667/1.1
[1] 14.33334
```

$s_A^2 = 1.1$이고 $s_B^2 = 15.76667$로 분산비는 약 14.333이 나옴

$H(0.95, 2, 5) = 7.15$보다 크므로 분산이 같다는 귀무가설 기각