HW02

이름 : 박유나 학번 2021-12659

Q1. a) $H_0: \sigma_1 = \sigma_2$ $H_a: \sigma_1 \neq \sigma_2$

p-value : 0.09 > 0.05 ⇒ H_0을 기각할 수 없다.

b) $H_0: \sigma_1 = \sigma_2$ $H_a: \sigma_1 \neq \sigma_2$

p-value : 0.0812 > 0.05 ⇒ H_0을 기각할 수 없다

c) $H_0: \sigma_1 = \sigma_2$ $H_a: \frac{\max(\sigma_1, \sigma_2)}{\max(\sigma_1, \sigma_2)} > 1$

p-value : 0.0485 < 0.05 ⇒ H_0을 기각할 수 있다. 즉, treatment 1과 treatment 2의 scale parameter는 다르다.

Q2. a) $T_{KW} = 3.62$

b) 5.991465 , 3.62 < 5.991465 ⇒ 유의수준 0.05에서 H_0을 기각할 수 없다

c) $Z_{JT} = 1.954601 > 1.645$ ⇒ 유의수준 0.05에서 H_0을 기각할 수 있다.

d) KW test와 JT test의 alternative hypothesis가 다른 때문이다. 그룹간의 순서가 있을 때 정보를 이용하기에 검정력의 차이가 있다.

Q3 a) $|\bar{X}_1 - \bar{X}_2| = 0.492$ $|\bar{X}_1 - \bar{X}_3| = 0.1784$ $|\bar{X}_2 - \bar{X}_3| = 1.286$

$|\bar{R}_1 - \bar{R}_2| = 1.4$ $|\bar{R}_1 - \bar{R}_3| = 5.2$ $|\bar{R}_2 - \bar{R}_3| = 3.8$

b) SSE = 13.840444

MSE = 1.153370

c) $q(0.05, 3, 12) = 3.773$

d) 1.810676. treatment 1, 2 / 2, 3 / 1, 3 간의 차이 있다

e) 6.62 treatment 1, 2 / 2, 3 / 1, 3 간의 차이 있다

Q4. p-value : 2.100e-05 < 0.05

⇒ 위의 모든 0.05에서 H_0을 기각할 수 있다.

즉, 가구 지출액들 간의 차이가 있다

HW02

2024-04-18

Q1

setting for computing

```
value <- c(8.0, 8.7, 8.8, 9.3, 9.7, 9.9, 10.0, 10.1, 10.6, 11.3, 6.1, 7.8, 8.1, 8.9, 9.0, 10.4,
10.5, 10.7, 10.8, 13.7)
label <- rep(c("treatment1", "treatment2"), each = 10)
df <- data.frame(value, label)


x <- with(df, value[label=="treatment1"])
y <- with(df, value[label=="treatment2"])

mean(x); mean(y)
```

```
## [1] 9.64
```

```
## [1] 9.6
```

```
m <- length(df$value[df$label=="treatment1"]); m
```

```
## [1] 10
```

```
n <- length(df$value[df$label=="treatment2"]); n
```

```
## [1] 10
```

```
N <- m + n

df <- df[order(df$value), ]
rank.st <- c(1, 4, 5, 8, 9, 12, 13, 16, 17, 20, 19, 18, 15, 14, 11, 10, 7, 6, 3, 2)
rank.ab <- c(seq(1:10), seq(from = 10, to = 1))
dev <- ifelse(df$label == "treatment1", abs(df$value - mean(x)), abs(df$value - mean(y)))

df[, "rank_st"] <- rank.st
df[, "rank_ab"] <- rank.ab
df[, "dev"] <- dev
```

observed data

```
D.observed <- with(df, mean(rank.st[label=="treatment1"])-mean(rank.st[label=="treatment2"]))
D2.observed <- with(df, mean(rank.ab[label=="treatment1"])-mean(rank.ab[label=="treatment2"]))
D3.observed <- with(df, mean(df$dev[label=="treatment1"]) / mean(df$dev[label=="treatment2"]))


D.observed; D2.observed; D3.observed
```

```
## [1] 4.6
```

```
## [1] 2.4
```

```
## [1] 0.4641975
```

permutation

```
B <- 10000
set.seed(0318)

# permutation for st test
D <- numeric()
for (i in 1:B) {
  rand.id <- sample(1:N, m, replace=FALSE)
  t1.st.permuted <- df$rank_st[rand.id]
  t2.st.permuted <- df$rank_st[-rand.id]
  D[i] <- mean(t1.st.permuted) - mean(t2.st.permuted)
}

# permutation for ab test

D2 <- numeric()
for (i in 1:B) {
  rand.id <- sample(1:N, m, replace=FALSE)
  t1.ab.permuted <- df$rank_ab[rand.id]
  t2.ab.permuted <- df$rank_ab[-rand.id]
  D2[i] <- mean(t1.ab.permuted) - mean(t2.ab.permuted)
}

# permutation for rmd test
D3 <- numeric()
for (i in 1:B) {
  rand.id <- sample(1:N, m, replace=FALSE)
  t1.rmd.permuted <- df$dev[rand.id]
  t2.rmd.permuted <- df$dev[-rand.id]
  D3[i] <- mean(t1.rmd.permuted)/mean(t2.rmd.permuted)
}
```

pvalue

```
pvalue.st <- 2*(sum(D >= D.observed))/B
pvalue.ab <- 2*(sum(D2 >= D2.observed))/B
pvalue.rmd <- 2*min( c(sum(D3 >= D3.observed)/B, sum(D3 <= D3.observed)/B) )

pvalue.st ; pvalue.ab ; pvalue.rmd
```

```
## [1] 0.09
```

```
## [1] 0.0812
```

```
## [1] 0.0458
```

Q2

a

```
value <- c(1.69, -0.90, 2.75, 0.51, 1.12, 2.33, 2.15, -0.39, 1.29, 2.25, 2.46, 2.43, 2.22, 2.9
6, 1.48)
label <- rep(c("t1", "t2", "t3"), each = 5)
df <- data.frame(value, label)
df[, "rank"] <- rank(df$value)

group.value <- aggregate(rank ~ label, data=df, length)[,2]
group.rank.mean <- aggregate(rank ~ label, data=df, mean)[,2]

N <- sum(group.value); K <- length(group.value)
KW.obs <- ( 12/(N*(N+1))) * sum( group.value * (group.rank.mean - (N+1)/2)^2 )
KW.obs
```

```
## [1] 3.62
```

b

```
qchisq(p = 0.05, df = 2, lower.tail = FALSE)
```

```
## [1] 5.991465
```

c

```
jt<-sum(sum(outer(df$value[label=="t1"], df$value[label=="t2"], "<")), sum(outer(df$value[label
=="t1"], df$value[label=="t3"], "<")),
sum(outer(df$value[label=="t2"], df$value[label=="t3"], "<")))



e.jt <- (15^2-3*5^2)/4
var.jt <- (15^2*(30+3) - 3*25*13)/72

z.jt <- (jt-e.jt)/sqrt(var.jt)
z.jt
```

```
## [1] 1.954601
```

3

a

```
means <- c(mean(df$value[label == "t1"]),mean(df$value[label == "t2"]), mean(df$value[label ==
"t3"]))
diff.means <- abs(outer(means, means, "-"))
diff.means
```

```
##       [,1]  [,2]  [,3]
## [1,] 0.000 0.492 1.276
## [2,] 0.492 0.000 0.784
## [3,] 1.276 0.784 0.000
```

```
df <- df %>% mutate(rank=rank(value))
rank.means <- c(mean(df$rank[label == "t1"]),mean(df$rank[label == "t2"]), mean(df$rank[label =
= "t3"]))
diff.rank.means <- abs(outer(rank.means, rank.means, "-"))
diff.rank.means
```

```
##      [,1] [,2] [,3]
## [1,]  0.0  1.4  5.2
## [2,]  1.4  0.0  3.8
## [3,]  5.2  3.8  0.0
```

b

```
model <- aov(value ~ label, data=df)

model
```

```
## Call:
##    aov(formula = value ~ label, data = df)
##
## Terms:
##                     label Residuals
## Sum of Squares   4.141493 13.840440
## Deg. of Freedom         2        12
##
## Residual standard error: 1.073951
## Estimated effects may be unbalanced
```

```
(SSE <- deviance(model))
```

```
## [1] 13.84044
```

```
(MSE <- deviance(model)/df.residual(model))
```

```
## [1] 1.15337
```

d

```
q <- 3.77
(qmse <- q*sqrt(MSE/5))
```

```
## [1] 1.810676
```

```
diff.means >= qmse
```

```
##       [,1]  [,2]  [,3]
## [1,] FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE
```

e

```
q2<- 3.31
(qrank <- q2*sqrt((15*16)/(12*5)))
```

```
## [1] 6.62
```

```
diff.rank.means >= qrank
```

```
##       [,1]  [,2]  [,3]
## [1,] FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE
```

4

```
scores <-
matrix(c(74, 88, 50,
         85, 87, 54,
         89, 93, 67,
         90, 96, 85,
         92, 99, 92,
         85, 92, 66,
         91, 97, 89,
         83, 93, 73,
         85, 94, 70,
         80, 90, 61,
         95, 100, 90),
       nrow = 11,
       byrow = TRUE,
       dimnames = list(1 : 11,
                       c("A", "B", "C")))
```

```
friedman.test(scores)
```

```
##
##  Friedman rank sum test
##
## data:  scores
## Friedman chi-squared = 21.535, df = 2, p-value = 2.107e-05
```