사회과학을위한데이터과학 과제5.

이름 : 김연주
학번 : 2021250461
학과 : 통계학과

1.

```
> #1
>
> library(dplyr)
>
> n1=276
> n2=468
> n3=87
>
> ybar1=10.4
> ybar2=7.4
> ybar3=8.3
>
> sd1=17.8
> sd2=13.6
> sd3=15.6
>
> set.seed(123)
>
> data=data.frame(
+   happiness = factor(rep(c("very", "pretty", "not very"), times=c(n1,
n2, n3))),
+   good_friends = c(rnorm(n1, ybar1,sd1), rnorm(n2, ybar2,sd2), rnorm
(n3, ybar3, sd3)))
>
>
> anova_results= aov(good_friends ~ happiness, data = data)
> summary(anova_results)
             Df Sum Sq Mean Sq F value Pr(>F)
happiness     2   1579   789.3   3.511 0.0303 *
Residuals   828 186151   224.8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> #(d)
> confint(anova_results)
                     2.5 %    97.5 %
(Intercept)      4.943925 11.254528
happinesspretty -4.018098  2.854081
happinessvery   -1.206158  6.031027
>
> #(e)
> confint(anova_results,level=1-0.05/3)
                    0.833 %  99.167 %
(Intercept)      4.243010 11.955443
happinesspretty -4.781387  3.617370
happinessvery   -2.009988  6.834857
```

(a)

source	Sum Sq	df	mean Sq	F	p-value
happiness	1579	2	789.3	3.511	0.0303
residuals	186151	828	224.8		

(b) H_0: population mean number of good friends is same among who reported different happiness

(vs) H_a: " is different "

F test statistic = $3.511 \sim F_{(2,828)}$

p-value = $0.0303 < 0.05$

$\Rightarrow$ reject the null.

population mean number of good friends is different among who reported different happiness

(c) sample size in each group is large $\rightarrow \bar{y}$ is approximately normal by CLT

random sample $\rightarrow$ not sure.

σ is same for each group $\rightarrow$ not sure.

(d) R 코드 참고

(e) R 코드 참고

(f) Bonferroni, 95% CI is wider.

2.

```
> #2
> student=read.table("C:\\Users\\kjhoe\\OneDrive\\바탕화~1-LAPTOP-VGOF9F
J7-25904429\\사회과학을위한데이터과학\\student.dat",header=T)
> student.1=student[,c(4,15)]
> head(student.1)
  traveltime G2
1          2 10
2          1  6
3          1 13
4          1  9
5          2 10
6          1 12
>
> traveltime_1 = subset(student.1, traveltime==1)
> traveltime_2 = subset(student.1, traveltime==2)
> traveltime_3 = subset(student.1, traveltime==3)
> traveltime_4 = subset(student.1, traveltime==4)
>
> n1 = length(traveltime_1)
> n2 = length(traveltime_2)
> n3 = length(traveltime_3)
> n4 = length(traveltime_4)
>
> ybar1 = mean(traveltime_1$G2)
> ybar2 = mean(traveltime_2$G2)
> ybar3 = mean(traveltime_3$G2)
> ybar4 = mean(traveltime_4$G2)
>
> s1 = sd(traveltime_1$G2)
> s2 = sd(traveltime_2$G2)
> s3 = sd(traveltime_3$G2)
> s4 = sd(traveltime_4$G2)
>
>
> data=data.frame(
+   traveltime = factor(rep(c(1,2,3,4), times=c(n1,n2,n3,n4))),
+   grade = c(rnorm(n1, ybar1,s1), rnorm(n2, ybar2,s2), rnorm(n3, ybar
3, s3),rnorm(n4, ybar4, s4)))
```

(a) $n_1 = 250$, $n_2 = 103$, $n_3 = 21$, $n_4 = 8$

$\bar{y}_1 = 11.012$, $\bar{y}_2 = 10.33981$, $\bar{y}_3 = 9.285711$, $\bar{y}_4 = 8$

$S_1 = 3.1360$, $S_2 = 3.0000$, $S_3 = 4.595$, $S_4 = 3.6645$

(b)

```
> anova_results= aov(grade~ traveltime, data = data)
> summary(anova_results)
             Df Sum Sq Mean Sq F value Pr(>F)
traveltime    3  71.55   23.85   1.148  0.432
Residuals     4  83.10   20.77
>
```

→ accept H_0 (population mean G2 is same among different traveltime)

(c) small sample size for n_4

randomization → not sure.

same σ → not sure.

(d)

```
> #(d)
> confint(anova_results)
                    2.5 %    97.5 %
(Intercept)    2.482695 20.379229
traveltime2 -11.244189 14.065332
traveltime3 -18.170895  7.138626
traveltime4 -17.482731  7.826790
```

(e)

```
>
> #(e)
> confint(anova_results,level=1-0.05/4)
                0.625 % 99.375 %
(Intercept)  -2.474839 25.33676
traveltime2 -18.255202 21.07634
traveltime3 -25.181908 14.14964
traveltime4 -24.493744 14.83780
```

(f)

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

Fit: aov(formula = grade ~ traveltime, data = data)

$traveltime
           diff       lwr      upr     p adj
2-1  1.4105714 -17.14398 19.96512 0.9882402
3-1 -5.5161347 -24.07069 13.03842 0.6529559
4-1 -4.8279709 -23.38252 13.72658 0.7290963
3-2 -6.9267061 -25.48126 11.62785 0.5042353
4-2 -6.2385423 -24.79309 12.31601 0.5745985
4-3  0.6881637 -17.86639 19.24272 0.9985690
```

(g) Bonferroni 95% is wider.

simultaneous CI.