숙제2

통계학전공 3학년 안효준 (5463312)

Problem 1.

$$F_0 = \frac{MSR}{MSE} = \frac{SSR/df(SSR)}{SSE/df(SSE)} = \frac{\frac{SSR}{SST}/df(SSR)}{\frac{SSE}{SST}/df(SSE)} = \frac{R^2/df(SSR)}{(1-R^2)/df(SSE)}, \ (0 \le R^2 \le 1)$$

따라서, R^2가 증가할수록 분모가 0에 가까워지므로 F_0값은 증가한다.

Problem 2.

$$E(y_{ij}) = E(a + bx_{ij} + \epsilon_{ij}) = E(a + bx_{ij}) + E(\epsilon_{ij}) = \boxed{a + bx_{ij}}$$

$$Var(y_{ij}) = Var(a + bx_{ij} + \epsilon_{ij}) = Var(\epsilon_{ij}) = \boxed{\sigma^2}$$

따라서, $y_{ij} \sim N(a + bx_{ij}, \sigma^2)$

Problem 3.

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix} = \begin{pmatrix} 1\,1\,0\,0\,0 \\ 1\,1\,0\,0\,0 \\ 1\,1\,0\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0 \\ 1\,0\,0\,1\,0 \\ 1\,0\,0\,1\,0 \\ 1\,0\,0\,0\,1 \\ 1\,0\,0\,0\,1 \\ 1\,0\,0\,0\,1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

$$Y \ = \ X \times \beta \ + \ \epsilon$$

모형식 : $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ (단, $i = 1,2,3,4, \quad j = 1,2,3$)

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

Problem 4.

$$(y_1 - \overline{y}_.) + (y_2 - \overline{y}_.) + \cdots + (y_5 - \overline{y}_.) = (y_1 + y_2 + \cdots + y_5) - 5\overline{y}_. = 5\overline{y}_. - 5\overline{y}_. = 0$$

$$(y_5 - \overline{y}_.) = -\left\{(y_1 - \overline{y}_.) + (y_2 - \overline{y}_.) + (y_3 - \overline{y}_.) + (y_4 - \overline{y}_.)\right\} = -(5 - 10 - 25 + 20) = \boxed{10}$$

Problem 5.

a)

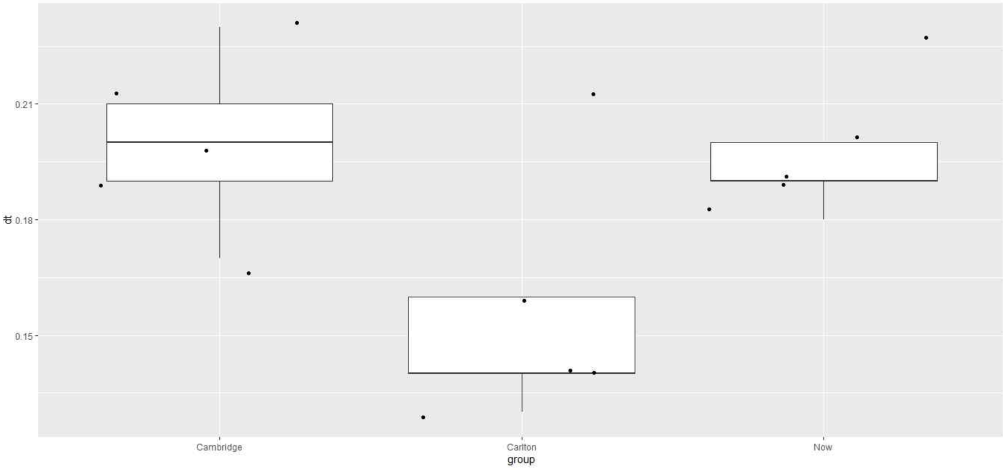
```
Carlton <- c(0.16, 0.14, 0.21, 0.14, 0.13)
Now <- c(0.19, 0.20, 0.23, 0.18, 0.19)
Cambridge <- c(0.21, 0.17, 0.19, 0.23, 0.20)

y <- c(Carlton, Now, Cambridge)
group <- rep(c('Carlton','Now','Cambridge'), each = 5)
```

```
df <- data.frame(y, group)

library(ggplot2)

ggplot(df, aes(x = group, y = dt)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter()
```



b)

$H_0 : \mu_1 = 0.5 \quad H_1 : \mu_1 \leq 0.5$

```
> t.test(Carlton, alternative = 'less', mu = 0.5)


        One Sample t-test

data:  Carlton
t = -23.968, df = 4, p-value = 8.987e-06
alternative hypothesis: true mean is less than 0.5
95 percent confidence interval:
      -Inf 0.1865978
sample estimates:
mean of x
    0.156
```

따라서 유의수준 0.05 하에서 귀무가설 기각 (Carlton 브랜드는 타르함량이 0.5보다 적다)

$H_0 : \mu_2 = 0.5 \quad H_1 : \mu_2 \leq 0.5$

```
> t.test(Now, alternative = 'less', mu = 0.5)


        One Sample t-test
```

```
data:  Now
t = -35.107, df = 4, p-value = 1.964e-06
alternative hypothesis: true mean is less than 0.5
95 percent confidence interval:
      -Inf 0.2163388
sample estimates:
mean of x
    0.198
```

따라서 유의수준 0.05 하에서 귀무가설 기각 (Now 브랜드는 타르함량이 0.5보다 적다)

$H_0 : \mu_3 = 0.5 \quad H_1 : \mu_3 \leq 0.5$

```
> t.test(Cambridge, alternative = 'less', mu = 0.5)

        One Sample t-test

data:  Cambridge
t = -30, df = 4, p-value = 3.676e-06
alternative hypothesis: true mean is less than 0.5
95 percent confidence interval:
      -Inf 0.2213185
sample estimates:
mean of x
      0.2
```

따라서 유의수준 0.05 하에서 귀무가설 기각 (Cambridge 브랜드는 타르함량이 0.5보다 적다)

c)

$H_0 : \tau_1 = \tau_2 = \tau_3 = 0 \quad H_1 : \exists \, \tau_i \;\; s.t. \;\; \tau_i \neq 0, \;\; (i = 1,2,3)$

```
> result <- aov(dt ~ group, data = df)
> summary(result)
            Df   Sum Sq   Mean Sq F value Pr(>F)
group        2 0.006173 0.0030867   4.874 0.0282 *
Residuals   12 0.007600 0.0006333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

따라서 유의수준 0.05 하에서 귀무가설 기각 (즉, 담배 브랜드에 따라 타르함량이 다르다)