Homework 2 (2023)

2019314199 김동환

1.

a)

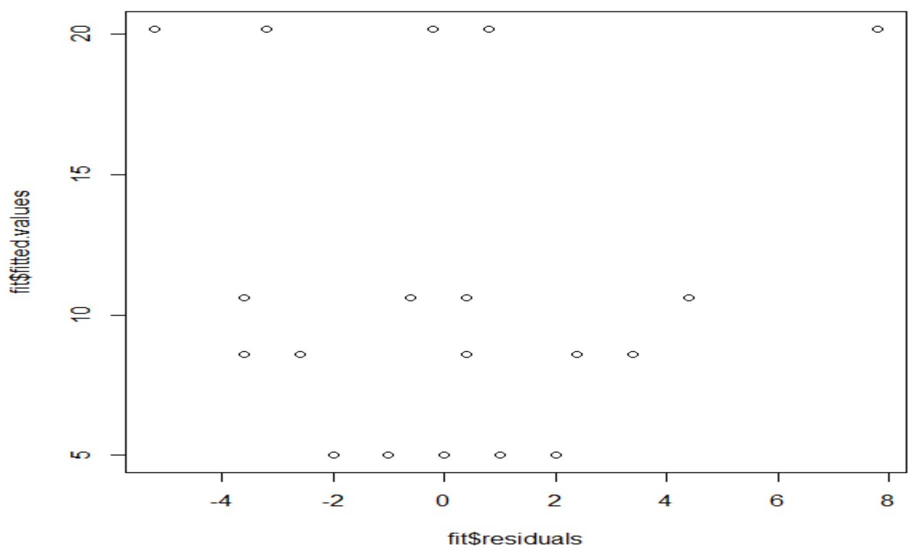
->Fit a one-way ANOVA and it's result

```
> x
 [1] 17 20 15 21 28  7 11 15 10 10 11  9  5 12  6  5  4  3  7  6
> y
 [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4
> fit<-aov(x~factor(y))
> anova(fit)
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value    Pr(>F)    
factor(y)  3  632.6  210.87  18.827 1.687e-05 ***
Residuals 16  179.2   11.20                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

해석: 그룹 A,B,C,D 간의 평균에 차이가 있다.

->plot residuals versus fitted values



This residuals plot tells me about constant variance assumption in the one-way ANOVA

->perform LeveneTest

```
> leveneTest(x~factor(y))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.9778 0.4278
      16
```

H0=Variances are constant

We cannot reject H0 under alpha=0.05

b)

Leven Test의 결과 등분산 가정을 기각하지 못함으로 굳이 변환을 안해도

된다고 생각하지만, 굳이 굳이 해야한다면 $\sqrt{x}$ 를 이용한다.

그 이유는 $P-value$ 가 더 작아져서 더욱 유의해지기 때문이다. 다른 변환방법인

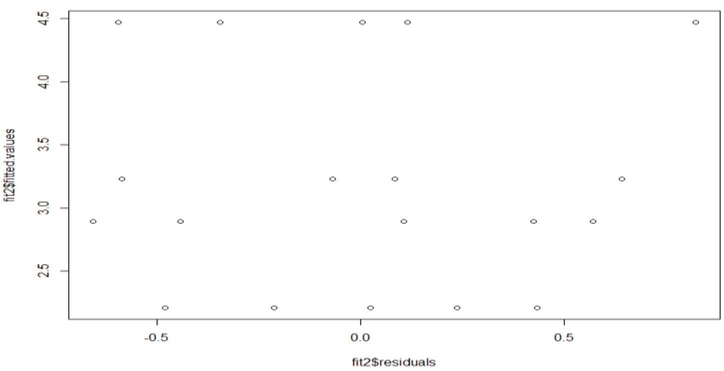
$\log(x)$, $1/x$, $1\sqrt{x}$ 들은 $P-value$가 변환 전보다 커진다.

c)

->Conduct the one-way ANOVA

```
> fit2<-aov(sqrt(x)~factor(y))
> anova(fit2)
Analysis of Variance Table

Response: sqrt(x)
          Df  Sum Sq Mean Sq F value    Pr(>F)
factor(y)  3 13.3946  4.4649  19.836 1.223e-05 ***
Residuals 16  3.6015  0.2251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



육안으로 residuals plot을 보았을 때는 큰 차이를 확인하기는 힘들다.

ANOVA 테이블에서는 $P-value$가 더 작아진 것을 확인 할 수 있다.

2.

a) $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

$\beta_j \sim N(0, \sigma_\beta^2)$ $\varepsilon_{ij} \sim N(0, \sigma^2)$

μ (overall mean) τ_i (ith treatment effect)

β_j (j th paitent effect) ε_{ij} (experimental error)

Patient are randomly selected so we consider it's effect as random (β_j)

Treatment effects are factor that researchers are interested in.

So we consider it's effect as fixed (τ_i)

b)

	Df	Sum Sq	Mean Sq	F	P
method	3	89083	29694	19.473	6.636e-05
block	4	1135104	283776	186.094	1.161e-10
Residuals	12	18299	1525		

$F = \frac{MS_{method}}{mse} = 19.473$ P-value 1.161e-10

→ 각 methods 별 평균에 차이가 있다.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: not H_0

$\alpha = 0.05$에서 H_0를 기각한다.

$F = \frac{MS\,block}{MSE} = 186.094$ P-value 6.636e-05

$\alpha = 0.05$에서 H_0를 기각한다.

$H_0: \mu_{p1} = \mu_{p2} = \mu_{p3} = \mu_{p4} = \mu_{p5}$

H_1: not H_0

→ blocking은 성공적으로 작동했다.

c) confidence interval for $\mu_1-\mu_4$

$$(\bar{y}_{1\cdot}-\bar{y}_{4\cdot}) \pm t_{\alpha/2,16}\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

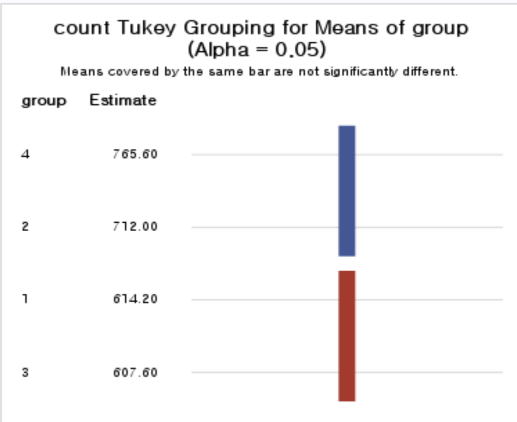
$$= -151.4 \pm 2.120\sqrt{1525\left(\frac{1}{5}+\frac{1}{5}\right)}$$

$$\rightarrow [-203.7601, -99.03986]$$

μ_4가 μ_1 보다 크고 신뢰구간은 위와 같다.

d)

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	1524.908
Critical Value of Studentized Range	4.19851
Minimum Significant Difference	73.322




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

Fit: aov(formula = y ~ method + blk)

$method
        diff        lwr       upr     p adj
m2-m1   97.8   24.47576 171.12424 0.0088574
m3-m1   -6.6  -79.92424  66.72424 0.9929651
m4-m1  151.4   78.07576 224.72424 0.0002578
m3-m2 -104.4 -177.72424 -31.07576 0.0055784
m4-m2   53.6  -19.72424 126.92424 0.1867736
m4-m3  158.0   84.67576 231.32424 0.0001735
```

method 2와 4는 평균에 유의한 차이가 없다.

method 1과 3은 평균에 유의한 차이가 없다.

3. $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \sim N(0, \sigma^2 + \sigma_\beta^2)$

$\bar{y}_{i.} = \mu + \tau_i + \bar{\beta} + \bar{\varepsilon}_{i.} \sim N(0. \frac{\sigma^2 + \sigma_\beta^2}{b})$

$\bar{y}_{.j} = \mu + \beta_j + \bar{\varepsilon}_{.j} \sim N(0, \frac{\sigma^2}{a} + \sigma_\beta^2)$

$\bar{y}_{..} = \mu + \bar{\beta}_. + \bar{\varepsilon}_{..} \sim N(0, \frac{\sigma^2}{ab} + \frac{\sigma_\beta^2}{b})$

$SS_{Trt} = b\sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 = b(\sum_{i=1}^{a} \bar{y}_{i.}^2 - 2\sum_{i=1}^{a} \bar{y}_{i.}\bar{y}_{..} + \sum_{i=1}^{a} \bar{y}_{..}^2)$

$= b(\sum_{i=1}^{a} \bar{y}_{i.}^2 - a\bar{y}_{..}^2)$

$E(SS_{Trt}) = bE(\sum_{i=1}^{a} \bar{y}_{i.}^2 - a\bar{y}_{..}^2) = b\sum_{i=1}^{a} (V(\bar{y}_{i.}) + E(\bar{y}_{i.})^2) - abE(V(\bar{y}_{..}) + E(\bar{y}_{..})^2)$

$= b\sum_{i=1}^{a} (\frac{\sigma^2 + \sigma_\beta^2}{b} + (\mu + \tau_i)^2) - ab(\frac{\sigma^2}{ab} + \frac{\sigma_\beta^2}{b} + \mu^2)$

$= a\sigma^2 + a\sigma_\beta^2 + b\sum_{i=1}^{a}(\mu + \tau_i)^2 - \sigma^2 - a\sigma_\beta^2 - ab\mu^2$

$= (a-1)\sigma^2 + b\sum_{i=1}^{a} \tau_i^2$

$MS_{Trt} = \frac{SS_{Trt}}{a-1}$ $\quad E(MS_{Trt}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$

E(MSblock)?

$$SS_{block} = a\sum_{j=1}^{b}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 = a\left(\sum_{j=1}^{b}\bar{y}_{\cdot j}^2 - b\bar{y}_{\cdot\cdot}^2\right)$$

$$E(SS_{block}) = a\left(\sum_{j=1}^{b}E(\bar{y}_{\cdot j}^2) - b\,E(\bar{y}_{\cdot\cdot}^2)\right) = a\sum_{j=1}^{b}\left(V(\bar{y}_{\cdot j}) + E(\bar{y}_{\cdot j})^2\right) - ab\left(V(\bar{y}_{\cdot\cdot}) + E(\bar{y}_{\cdot\cdot})^2\right)$$

$$= a\sum_{j=1}^{b}\left(\frac{\sigma^2}{a} + \sigma_\beta^2 + \mu^2\right) - ab\left(\frac{\sigma^2}{ab} + \frac{\sigma_\beta^2}{b} + \mu^2\right) = b\sigma^2 + ab\sigma_\beta^2 - \sigma^2 - a\sigma_\beta^2$$

$$= (b-1)\sigma^2 + a(b-1)\sigma_\beta^2 = (b-1)(\sigma^2 + a\sigma_\beta^2)$$

$$MS_{block} = \frac{SS_{block}}{b-1} \qquad E(MS_{block}) = \sigma^2 + a\sigma_\beta^2$$

E(MSE)?

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot j} + \bar{\varepsilon}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\left((\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2 - 2(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}) + (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2\right)$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2 - a\sum_{j=1}^{b}(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2$$

$$E(SS_E) = E\left(\sum_{i=1}^{a}\sum_{j=1}^{b}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2\right) - aE\left(\sum_{j=1}^{b}(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2\right)$$

$$\frac{\sum_{j=1}^{b}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2}{\sigma^2} \sim \chi(b-1) \rightarrow E\left(\sum_{j=1}^{b}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2\right) = \sigma^2(b-1)$$

$$\frac{\sum_{j=1}^{b}(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2}{\sigma^2/a} \sim \chi(b-1) \rightarrow E\left(\sum_{j=1}^{b}(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2\right) = \frac{\sigma^2}{a}(b-1)$$

$$= a(b-1)\sigma^2 - (b-1)\sigma^2 = (a-1)(b-1)\sigma^2$$

$$E(MS_E) = E\left(\frac{SS_E}{(a-1)(b-1)}\right) = \sigma^2$$