3.1.

2017100057
통계학과 이영노

(1) $RR = \dfrac{\text{Treatment group risk}}{\text{Control group risk}}$ $\quad OR = \dfrac{\text{Treatment group odds}}{\text{Control group odds}}$

$\therefore RR = \dfrac{5/21}{8/82} = 2.440476.\quad OR = \dfrac{5/13}{8/13} \Big/ \dfrac{21/103}{82/103} = 2.440476$

(2). RR의 $100(1-\alpha)\%$ C.I. : $\{RR\cdot\exp(-z_{\frac{\alpha}{2}}\cdot\sqrt{v_1}),\ RR\cdot\exp(z_{\frac{\alpha}{2}}\cdot\sqrt{v_1})\}$

OR 〃 〃 : $\{OR\cdot\exp(-z_{\frac{\alpha}{2}}\cdot\sqrt{v_2}),\ OR\cdot\exp(z_{\frac{\alpha}{2}}\cdot\sqrt{v_2})\}$,

where $v_1 = \dfrac{1-n_{11}/n_{1+}}{n_{11}} + \dfrac{1-n_{21}/n_{2+}}{n_{21}} = \dfrac{1-5/26}{5} + \dfrac{1-8/90}{8} = 0.2754.$

$v_2 = \dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} + \dfrac{1}{n_{21}} + \dfrac{1}{n_{22}} = \dfrac{1}{5} + \dfrac{1}{8} + \dfrac{1}{21} + \dfrac{1}{82} = 0.3848$

$z_{\frac{\alpha}{2}} = 1.96$ when $\alpha = 0.05$.

RR의 $100(1-\alpha)\%$ C.I. : $\{2.440476\cdot e^{\pm 1.96\cdot\sqrt{0.2754}}\} = \{0.896,\ 6.826\}$.

OR 〃 〃 : $\{$ 〃 $\cdot e^{\pm 1.96\sqrt{0.3848}}\} = \{0.724,\ 8.232\}$.

```r
## 3.1
```{r}
rr=(5/21)/(8/82)
or=((5/13)/(8/13))/((21/103)/(82/103))
v1=((1-5/26)/5)+((1-8/90)/8)
v2=1/5+1/8+1/21+1/82
RRCI95 = c(rr*exp(-qnorm(0.975)*sqrt(v1)),rr*exp(qnorm(0.975)*sqrt(v1)))
ORCI95 = c(or*exp(-qnorm(0.975)*sqrt(v2)),or*exp(qnorm(0.975)*sqrt(v2)))

c(rr,or)
c(v1,v2)
c("RR 95% Confidence Interval is,",RRCI95)
c("OR 95% Confidence Interval is,",ORCI95)
```
```

```
[1] 2.440476 2.440476
[1] 0.2754274 0.3848142
[1] "RR 95% Confidence Interval is," "0.872477069754962"            "6.8264533736731"
[1] "OR 95% Confidence Interval is," "0.723511413977944"            "8.23196969835605"
```

(3) (1). RR : 위험요인이 있을때 산후우울증 발병률이 위험요인이 없을때 산후우울증 발병률 보다 2.440476 배 많다.

OR : 산후우울증이 발병한 집단에서 위험요인이 있었던 표본이, 산후우울증이 발병하지 않았던 집단보다 2.440476 배 많다.

(2) RR, OR의 신뢰구간 둘다 1을 포함하므로, RR, OR 둘다 1과 유의한 차이가 있다고 말하기 어렵다. 즉, 위험요인 유무가 산후우울증 발병위험을 높인다고 말할 통계적 근거가 부족하다.

#3.3

H_0 : 세 그룹간 재발률이 동일하다. $(p_{1j} = p_{2j} = p_{3j}\ ,\ j=1,2)$

H_α : 세 그룹간 재발률이 동일하지 않다.

under H_0, 기대도수 E_{ij} for ith group, jth result. is $n_{i+} \cdot n_{+j} / n$

under H_0, $\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{2} (n_{ij} - E_{ij})^2 / E_{ij} \sim \chi^2_{df=(3-1)(2-1)=2}$.

table for E_{ij} under H_0 is,

61·75/125 = 36.6	61·50/125 = 24.4	61
33·75/125 = 19.8	33·50/125 = 13.2	33
31·75/125 = 18.6	31·50/125 = 12.4	31
75	50	125

table for $(n_{ij} - E_{ij})^2 / E_{ij}$ under H_0 is,

$(49-36.6)^2/36.6$ = 4.201093	$(12-24.4)^2/24.4$ = 6.301639	61
$(24-19.8)^2/19.8$ = 0.890909	$(9-13.2)^2/13.2$ = 1.336364	33
$(2-18.6)^2/18.6$ = 14.81505	$(29-12.4)^2/12.4$ = 22.22258	31
75	50	125

$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{2} (n_{ij} - E_{ij})^2 / E_{ij} = 49.76764 . \sim \chi^2_{df=2}$.

qchisq(0.95, df=2) = 5.9915 $< \chi^2 = 49.76764$

so we reject H_0. (under $\alpha = 0.05$)

3.3

```{r}
reject=c(49,12); recommend=c(24,9); surgery=c(2,29);
df3.3 = rbind(reject, recommend, surgery)
colnames(df3.3)=c("Yes","No")

df3.3
chisq.test(df3.3,correct=FALSE)
```

```
        Pearson's Chi-squared test

data:  df3.3
X-squared = 49.768, df = 2, p-value = 1.56e-11
```

which means "세 그룹간 재발률이 동일하지 않다."

#3.5

H_0: 방사선 치료 유무와 후두암 치료 결과 간에는 아무런 연관이 없다.

H_α: 수술시 후두암 치료율이 방사선 치료시 후두암 치료율 보다 높다. (단측검정).

21	2	23
15	3	18
36	5	41

table for E_{ij} under H_0.

$36 \cdot 23/41 = 20.\ldots$	$5 \cdot 23/41 = 2.80\ldots$
$36 \cdot 18/41 = 15.8\ldots$	$5 \cdot 18/41 = 2.19\ldots$

⇒ 기대도수가 5 미만인 cell이 50% 이므로 χ^2 통계량의 분모가 작아져 통계량을 과대추정 할 우려가 있으므로,

Fishers exact Test를 한다.

$$p(n_{11}=21) = \binom{36}{21}\binom{5}{2}\Big/\binom{41}{23}$$

$$p(n_{11}=22) = \binom{36}{22}\binom{5}{1}\Big/\binom{41}{23}$$

$$p(n_{11}=23) = \binom{36}{23}\binom{5}{0}\Big/\binom{41}{23}$$

$p(n_{11}=21)+p(n_{11}=22)+p(n_{11}=23) = p.value.$

$= 0.3808337. > 0.05.$

since pvalue $> \alpha = 0.05$, we do not reject H_0.

```
## 3.5
```{r}
choose(36,21)*choose(5,2)/choose(41,23)+choose(36,22)*choose(5,1)/choose(41,23)+
  choose(36,23)*choose(5,0)/choose(41,23)

surgery=c(21,2); radio=c(15,3);
df3.5=rbind(surgery,radio)
colnames(df3.5)=c("Cured","Not Cured")

df3.5
fisher.test(df3.5,alternative="greater",conf.int=FALSE)
```
```

```
	Fisher's Exact Test for Count Data

data:  df3.5
p-value = 0.3808
alternative hypothesis: true odds ratio is greater than 1
sample estimates:
odds ratio
  2.061731
```

which means that

방사선 치료 유무와 후두암 치료 결과 간에는 아무런 연관이 없다.

#3.7

Comparing proportion of paired sample. → McNemar's Test.

H_0: 첫번째 투표시 찬성의 비율이 두번째 투표시 찬성비율과 동일하다.

H_α: 첫번째 투표와 두번째 투표시 찬성의 비율이 다르다.

under H_0,

$$Q = (n_{12}-n_{21})^2/(n_{12}+n_{21}) = (150-86)^2/(150+86) = 17.3559. \dot{\sim} \chi^2_{df=1}$$

$$\chi^2_{\alpha=0.05,\,df=1} = 3.84 < Q = 17.3559.$$

so we reject H_0 (under $\alpha = 0.05$)

```
## 3.7
```{r}
a=c(794,150); b=c(86,570);
df3.7=rbind(a,b)
df3.7
mcnemar.test(df3.7,correct=FALSE)
```
```

```
	McNemar's Chi-squared test

data:  df3.7
McNemar's chi-squared = 17.356, df = 1, p-value = 3.099e-05
```

which means that

첫번째 투표와 두번째 투표시 찬성의 비율이 다르다.

3.8

Suppose we look at table of gender and survival with respect to kth strata

kth strata's table of gender and survival.

		Survival Alive	Dead	
gender	F	n_{k11}	n_{k12}	n_{k1+}
	M	n_{k21}	n_{k22}	n_{k2+}
		n_{k+1}	n_{k+2}	n_k

H_0 : 남자의 사망율은 여자의 사망율과 동일하다.

H_a : 남자의 사망율은 여자의 사망율과 동일하지 않다.

under H_0,

$$Q_{CMH} = \frac{\left[\sum_{k=1}^{4} n_{k11} - \sum_{k=1}^{4} E(n_{k11} \mid H_0)\right]^2}{\sum_{k=1}^{4} Var(n_{k11} \mid H_0)} \dot{\sim} \chi^2_{df=1} \quad \text{where,}$$

$$E(n_{k11} \mid H_0) = \frac{n_{k1+} \cdot n_{k+1}}{n_k}, \quad Var(n_{k11} \mid H_0) = \frac{n_{k1+} \cdot n_{k2+} \cdot n_{k+1} \cdot n_{k+2}}{n_k^2 (n_k - 1)}$$

$$E(n_{k11} \mid H_0) = \begin{cases} 23 \cdot 2 / 33 & k=1 \\ 42 \cdot 2 / 60 & k=2 \\ 39 \cdot 6 / 49 & k=3 \\ 33 \cdot 17 / 37 & k=4 \end{cases} \quad Var(n_{k11} \mid H_0) = \begin{cases} 23 \cdot 10 \cdot 2 \cdot 31 / 33^2 \cdot 32 & k=1 \\ 42 \cdot 18 \cdot 2 \cdot 58 / 60^2 \cdot 59 & k=2 \\ 39 \cdot 10 \cdot 6 \cdot 43 / 49^2 \cdot 48 & k=3 \\ 33 \cdot 4 \cdot 17 \cdot 20 / 37^2 \cdot 36 & k=4 \end{cases}$$

$$\rightarrow \sum_{k=1}^{4} E(n_{k11} \mid H_0) = 22.73161 \quad \sum_{k=1}^{4} Var(n_{k11} \mid H_0) = 2.605801.$$

$$\rightarrow Q_{CMH} = (27 - 22.73161)^2 / 2.605801 = 6.991767. \dot{\sim} \chi^2_{df=1}$$

$$\chi^2_{\alpha=0.05, df=1} = 3.84 < Q_{CMH} = 6.991767$$

$\therefore$ reject H_0 which means that

남자의 사망율은 여자의 생존율과 동일하지 않다.

3.8
```{r}
# number of category is 16
strata = factor(c("심각하지 않음","심각하지 않음","심각하지 않음","심각하지 않음",
          "약간심각","약간심각","약간심각","약간심각",
          "심각","심각","심각","심각","매우심각","매우심각","매우심각","매우심각"))
gender = factor(c("M","M","F","F","M","M","F","F","M","M","F","F","M","M","F","F"))
vital = factor(c("Dead","Alive","Dead","Alive","Dead","Alive","Dead","Alive",
         "Dead","Alive","Dead","Alive","Dead","Alive","Dead","Alive"))
count = c(2,21,0,10,2,40,0,18,6,33,0,10,17,16,0,4)
strata=as.factor(strata); gender=as.factor(gender); vital=as.factor(vital)
vital=relevel(vital,"Dead"); gender=relevel(gender,"M")


dat3.8 = data.frame(strata,gender,vital,count)

# 3차원 이상의 분할표 사용시 `ftable()`
tab3.8 = xtabs(count ~ gender+vital+strata, data=dat3.8)
mantelhaen.test(tab3.8,correct=FALSE)
```

```
        Mantel-Haenszel chi-squared test without continuity correction

data:  tab3.8
Mantel-Haenszel X-squared = 6.9918, df = 1, p-value = 0.008189
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 NaN NaN
sample estimates:
common odds ratio
              Inf
```

3.9

(1) Sensitivity $= \dfrac{302}{481} = P(T+|D+)$ Specificity $= \dfrac{372}{452} = P(T-|D-)$.

(2) 유병률 $= 0.1 = P(D+)$.

$$PVP \equiv P(D+|T+) = \frac{P(D+, T+)}{P(T+)} = \frac{P(T+|D+)\cdot P(D+)}{P(T+|D+)\cdot P(D+) + P(T+|D-)\cdot P(D-)}$$

$$= \frac{\text{Sensitivity} \times \text{유병률}}{(\text{Sensitivity} \times \text{유병률}) + (1-\text{Specificity})(1-\text{유병률})}$$

$$= \frac{302}{481}\cdot 0.1 \Big/ \left(\frac{302}{481}\cdot 0.1\right) + \left(1-\frac{372}{452}\right)(1-0.1)$$

$$= 0.28272$$

$$PVN \equiv P(D-|T-) = \frac{P(D-, T-)}{P(T-)} = \frac{P(T-|D-)P(D-)}{P(T-|D+)P(D+) + P(T-|D-)P(D-)}$$

$$= \frac{\text{Specificity} \times (1-\text{유병률})}{(1-\text{Sensitivity})\,\text{유병률} + \text{Specificity}\cdot(1-\text{유병률})}$$

$$= \frac{\frac{372}{452}\cdot(1-0.1)}{\left(1-\frac{302}{481}\right)\cdot 0.1 + \frac{372}{452}\cdot(1-0.1)}$$

$$= 0.9521621 .$$

```
## 3.9
```{r}
diagnosis = factor(c("Positive","Negative","Positive","Negative"));
disease = factor(c("Yes","Yes","No","No"));
freq = c(302,179,80,372);
diagnosis=relevel(diagnosis,"Positive"); disease=relevel(disease,"Yes")
tab3.9 = xtabs(freq~diagnosis+disease)

margin1=margin.table(tab3.9, margin=1) # margin about disease(result is row)
margin2=margin.table(tab3.9, margin=2) # margin about diagnosis(result is columns)

#민감도,특이도
sensitivity = tab3.9[1,1]/margin2[1]
print(sensitivity)
specificity = tab3.9[2,2]/margin2[2]
print(specificity)

#PVP
pvp = (302/481*0.1)/((302/481*0.1)+(1-372/452)*0.9)
print(pvp)
#PVN
pvn = (372/452*0.9)/((372/452*0.9)+(1-302/481)*0.1)
print(pvn)
```
```

```
      Yes
0.6278586
       No
0.8230088
[1] 0.28272
[1] 0.9521621
```

2.2

(1) H_0 : 림프구세포의 모분산(σ_1^2)이 종양세포의 모분산(σ_2^2)과 같다.

$(\sigma_1^2/\sigma_2^2 = 1)$.

$H_a : \sigma_1^2/\sigma_2^2 \neq 1$.

under H_0, $F = \dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} = S_1^2/S_2^2$.

$$S_1^2 = \frac{1}{n_1-1}\sum_{\forall i}(X_{1i} - \overline{X_1})^2 = 2.546667$$

$$S_2^2 = \frac{1}{n_2-1}\sum_{\forall i}(X_{2i} - \overline{X_2})^2 = 7.987122$$

$$F = \frac{S_1^2}{S_2^2} = 0.3188466$$

$$F_{1-0.05/2\,(df_1=39,\ df_2=39)} = 0.5288993 > F = \frac{S_1^2}{S_2^2} = 0.3188$$

So we reject the H_0 which means that $\sigma_1^2 \neq \sigma_2^2$.

H_0 : 림프구세포의 크기평균(μ_1)이 종양세포의 크기평균(μ_2)과 같다.

$(\mu_1 = \mu_2)$.

$H_a : \mu_1 < \mu_2$.

Comparing mean of two independent groups.

모분산이 다른 경우.

under H_0,

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t_v \quad \text{where } v \text{ is Satterthwaite df.}$$

$$w_1 = \frac{S_1^2}{n_1} = \frac{2.5466}{40} = 0.063667$$

$$w_2 = \frac{S_2^2}{n_2} = \frac{7.987122}{40} = 0.199678$$

$$v = \frac{(w_1 + w_2)^2}{\frac{w_1^2}{n_1-1} + \frac{w_2^2}{n_2-1}} = 61.5749 \approx 62$$

$$t = \frac{(6.95 - 17.6575) - 0}{\sqrt{\frac{2.546667}{40} + \frac{7.987122}{40}}} = \frac{-10.7075}{0.51317} = -20.86536 \sim t_{df=v=62}$$

$$-t_{0.95,\, df=62} = -1.669804 > t_v = -20.8653\ldots$$

So we reject the H_0, which means that

림프구 세포의 크기(평균)이 중앙세포의 크기(평균)보다 작다. $(\mu_1 < \mu_2)$.

(2) 표본들이 각각 정규분포인 모집단으로부터 추출되었다.

```r
# f test for equal variance
var.test(X,Y,alternative="two.sided")
# two sample t test with different variance
t.test(X,Y,var.equal = FALSE, alternative="less")
```

```
        F test to compare two variances

data:  X and Y
F = 0.31885, num df = 39, denom df = 39, p-value = 0.0005464
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1686378 0.6028493
sample estimates:
ratio of variances
         0.3188466


        Welch Two Sample t-test

data:  X and Y
t = -20.865, df = 61.575, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -9.850515
sample estimates:
mean of x mean of y
   6.9500   17.6575
```

#2.4

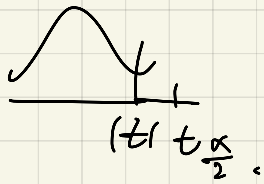
(1) H_0: 사전 점수와 사후점수의 차이의 모평균 μ_d가 0이다. ($\mu_d=0$)

H_a: 〃 〃 0이 아니다. ($\mu_d \neq 0$).

under H_0,

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{(-2.09091) - 0}{4.276787/\sqrt{11}} = -1.621488 \sim t_{df=11-1=10}.$$

$$t_{\frac{0.05}{2}, df=10} = 2.228139 > |t| = 1.621488$$



So we do not reject H_0, which means

사전점수와 사후점수의 차이의 모평균 μ_d가 0이다. 차이가 없다.

```
t.test(prior-post)
```

```
	One Sample t-test

data:  prior - post
t = -1.6215, df = 10, p-value = 0.136
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -4.9640939  0.7822757
sample estimates:
mean of x
-2.090909
```

(2) for μ_d, 95% CI is $\{\bar{d} \pm t_{0.025, df=10} \cdot \frac{s_d}{\sqrt{n}}\}$

$$= \{-2.09091 \pm 2.228139 \cdot \frac{4.276787}{\sqrt{11}}\} = \{-4.964095, 0.7822748\}$$

(3) (2)에서 나온 95% CI가 (1)에서 구한 $\mu_d=0$을 포함하는 범위이다. 따라서, 신뢰구간이 검정결과를 뒷받침해준다.

(결과에 일관성이 있다.)

#2.7

⟨Two-way ANOVA⟩

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}. \quad (i=1\cdots4,\ j=1\cdots5,\ k=1)$$

$$(\varepsilon_{ijk} \sim N(0, \sigma^2))$$

$$\left(\sum_i \alpha_i = 0\ \ \sum_j \beta_j = 0\ \ \sum(\alpha\beta)_{ij} = 0\right)$$

$H_0 : (\alpha\beta)_{ij} = 0$. H_α : not all $(\alpha\beta)_{ij}$ is 0.

$$SS_{A\times B} = n \sum_{\forall i}\sum_{\forall j} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = 9.28_{//}$$

cell 하나당 sample 수가 1이니, $H_0 : (\alpha\beta)_{ij} = 0$ 가정하에서
SSE를 구할 수 없다. 따라서 교호작용이 없는 모형을 가정하
고 ANOVA를 해야 한다

$$SS_A = n \cdot J \cdot \sum_{\forall i} (\bar{y}_{i..} - \bar{y}_{...})^2 = 5\left[\begin{array}{l}(8.16 - 6.755)^2 + (5.38 - 6.755)^2 + \\ (5.84 - 6.755)^2 + (7.64 - 6.755)^2\end{array}\right]$$

$$= 5 \cdot (5.4851) = 27.4255$$

$$MS_A = SS_A/(I-1) = 27.4255/3 = 9.141833$$

$$SS_B = n \cdot I \sum_{\forall j} (\bar{y}_{.j.} - \bar{y}_{...})^2 = 62.65.$$

$$MS_B = SS_B/(J-1) = 62.65/4 = 15.662$$

$$SS_e = SST - SS_A - SS_B = \sum\sum(y_{ij} - \bar{y})^2 - J\sum(\bar{y}_{i.} - \bar{y})^2 - I\sum(\bar{y}_{.j} - \bar{y})^2$$

$$= 99.36 - 27.43 - 62.65 = 9.28$$

$$MS_e = 9.28/4 \cdot 3 = 0.773$$

All preparation is set. Now, let's do Hypothesis Testing

(1) $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ vs. H_a: not all α_i is 0.

$$F_A = MS_A / MS_E = 15.662 / 0.773 = 20.26 \sim \chi^2_{(df1=3,\ df2=12)}$$

$$F_{\alpha=0.05,\ 3,12} = 3.490295 < F_A = 20.26$$

So we reject H_0, which means

not all α_i is 0. 식이요법에 따라 체중감량에 차이가 있음.

(2) $H_0: \beta_1 = \cdots = \beta_5 = 0$ vs. H_a: not all β_i is 0

Two-way로 가야 이게 왜 붙음? ANOVA 위도 다시.

$$F_B = MS_B / MS_E = 9.141822 / 0.773 = 11.82 \sim \chi^2_{(df1=4,\ df2=12)}.$$

$$F_{\alpha=0.05,\ 4,12} = 7.6163 < F_B = 11.82$$

So we reject H_0 which means

not all β_i is 0. 운동방법에 따라 체중감량에 차이가 있음.

```
model.aov3 <- aov(Y~exercise+food)
summary(model.aov3)

```
```

```
          Df Sum Sq Mean Sq F value   Pr(>F)    
exercise   4  62.65  15.662   20.26 2.87e-05 ***
food       3  27.43   9.142   11.82 0.000677 ***
Residuals 12   9.28   0.773                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
```

(3)

```
ScheffeTest(model.aov3)
```

```
  Posthoc multiple comparisons of means: Scheffe Test
    95% family-wise confidence level

$exercise
      diff      lwr.ci      upr.ci    pval
B-A  1.525 -1.2826574  4.33265744 0.56267
C-A  0.200 -2.6076574  3.00765744 1.00000
D-A -2.825 -5.6326574 -0.01734256 0.04814 *
E-A  2.375 -0.4326574  5.18265744 0.12592
C-B -1.325 -4.1326574  1.48265744 0.70973
D-B -4.350 -7.1576574 -1.54234256 0.00183 **
E-B  0.850 -1.9576574  3.65765744 0.95553
D-C -3.025 -5.8326574 -0.21734256 0.03106 *
E-C  2.175 -0.6326574  4.98265744 0.18857
E-D  5.200  2.3923426  8.00765744 0.00035 ***

$food
     diff     lwr.ci     upr.ci   pval
2-1 -2.78 -5.2912452 -0.2687548 0.0259 *
3-1 -2.32 -4.8312452  0.1912452 0.0794 .
4-1 -0.52 -3.0312452  1.9912452 0.9947
3-2  0.46 -2.0512452  2.9712452 0.9975
4-2  2.26 -0.2512452  4.7712452 0.0916 .
4-3  1.80 -0.7112452  4.3112452 0.2572

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

감소된 체중의 평균 차이가 통계적으로 유의미한 건 D∼A,B,C,E , 1∼2 조합이었고, 해당 조합에서의 감소된 체중의 평균의 차이는, 위에 해당하는 diff와 같다.

#2.8

Two way ANOVA.

k=1 이라서 residual의 자유도가 없기때문에 교호작용에 대한 $MS_{A\times B}$를 구하지 못하여 $(\alpha\beta)_{ij}=0$ 를 검정하지 못한다.

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \left(i=1\sim5,\ j=1\sim3 \quad \varepsilon_{ij} \sim N(0,\sigma^2),\ \sum\alpha_i=0,\ \sum\beta_j=0 \right)$$

$$SS_A = J\cdot\sum(\bar{y}_{i\cdot}-\bar{y})^2 = 3\cdot(129.7) = 389.1$$

$$MS_A = SS_A/(I-1) = 389.1/4 = 97.27$$

$$SS_B = I\cdot\sum(\bar{y}_{\cdot j}-\bar{y})^2 = 5\cdot 42.98 = 214.9$$

$$MS_B = SS_B/(J-1) = 214.9/2 = 107.47$$

$$SS_E = SST - SS_A - SS_B = 643.7 - 389.1 - 214.9 = 39.7$$

$$MS_E = SS_E/(I-1)(J-1) = 39.7/4\cdot2 = 4.97$$

(1) $H_0: \alpha_1 = \cdots = \alpha_5 = 0$ $\quad H_a$: not all α_i is 0.

$$F_A = MS_A / MS_E = 97.27 / 39.7 = 19.58 \sim F_{(df1=4,\ df2=8)}$$

$$F_{0.05,4,8} = 3.837853 < F_A = 19.58$$

So we reject H_0, which means

not all α_i is 0. 방법의 평균간에 차이가 존재한다.

(2) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$. $\quad H_a$: not all β_i is 0.

$$F_B = MS_B / MS_E = 107.47 / 4.97 = 21.62374 \sim F_{(df1=2,\ df2=8)}.$$

$$F_{0.05,2,8} = 4.458997 < F_B = 21.62374$$

So we reject H_0, which means,

not all β_i is 0. 흥미 동기부여 수준에 따라 차이가 있음.

(3)

```
library(multcompView)
TukeyHSD(model.aov3)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Y ~ method + motivation)

$method
          diff         lwr       upr     p adj
2-1   9.666667   3.3802451 15.953088 0.0046864
3-1   2.000000  -4.2864215  8.286422 0.8028131
4-1  14.000000   7.7135785 20.286422 0.0003971
5-1   5.333333  -0.9530882 11.619755 0.1022977
3-2  -7.666667 -13.9530882 -1.380245 0.0182027
4-2   4.333333  -1.9530882 10.619755 0.2138916
5-2  -4.333333 -10.6197549  1.953088 0.2138916
4-3  12.000000   5.7135785 18.286422 0.0011509
5-3   3.333333  -2.9530882  9.619755 0.4190608
5-4  -8.666667 -14.9530882 -2.380245 0.0090740

$motivation
                diff        lwr        upr     p adj
No-Low          -9.2 -13.227544 -5.1724562 0.0004716
Very Low-Low    -3.6  -7.627544  0.4275438 0.0781255
Very Low-No      5.6   1.572456  9.6275438 0.0101623
```

1~2,4

3~2,4

4~5

없음~낮음, 매우낮음.

간의 관계가 diff 만큼 차이나고, 이들이 통계적으로 유의미하다.