Question 2

(Analysis of binary: see Example 6.3) Table 6.3, from Crowder (1978), shows the

results of a 2 \_ 2 factorial experiment on seed variety and type of root extract.

The outcome yi is the number of seeds that germinated out of ni planted seeds.

> rm(list=ls())

> #########################################################################

> #########################################################################

> #question2

> #########################################################################

> #########################################################################

>

> getwd()

[1] "D:/강재훈/연세대학교/석사/2019\_1\_석사\_2학기/수업/중응통/HW/HW2/question\_2"

> setwd("D:/강재훈/연세대학교/석사/2019\_1\_석사\_2학기/수업/중응통/HW/HW2/question\_2")

>

> df2 = read.table('seed.dat')

> mat2 = as.matrix(df2)

>

> #number of seeds that germinated

> y\_i = mat2[,1]

> #number of planted seeds

> n\_i = mat2[,2]

>

> #(a) Calculate yi=ni's and compare their averages for Seed and Extract.

> seed\_A = 1:11

> seed\_B = 12:21

> extr\_1 = c(1:5,12:16)

> extr\_2 = c(6:11,17:21)

>

> seed\_extr\_A1 = 1:5

> seed\_extr\_A2 = 6:11

> seed\_extr\_B1 = 12:16

> seed\_extr\_B2 = 17:21

>

> #calculate y\_i/n\_i

> avgs = y\_i/n\_i

>

> #average of seed A

> sum(y\_i[seed\_A])/sum(n\_i[seed\_A])

[1] 0.5291005

> #average of seed B

> sum(y\_i[seed\_B])/sum(n\_i[seed\_B])

[1] 0.469697

> #average of Extract1

> sum(y\_i[extr\_1])/sum(n\_i[extr\_1])

[1] 0.3746835

> #average of Extract2

> sum(y\_i[extr\_2])/sum(n\_i[extr\_2])

[1] 0.6330275

>

> #average of seed A & extract 1

> sum(y\_i[seed\_extr\_A1])/sum(n\_i[seed\_extr\_A1])

[1] 0.3639706

> #average of seed A & extract 2

> sum(y\_i[seed\_extr\_A2])/sum(n\_i[seed\_extr\_A2])

[1] 0.6813559

> #average of seed B & extract 1

> sum(y\_i[seed\_extr\_B1])/sum(n\_i[seed\_extr\_B1])

[1] 0.398374

> #average of seed B & extract 2

> sum(y\_i[seed\_extr\_B2])/sum(n\_i[seed\_extr\_B2])

[1] 0.5319149

>

> #(b) Gabdol thinks that we can apply the two-way ANOVA for yi=ni's because

yi=ni's are approximately normally distributed. Do it. Any comment to the result?

> #Two way ANOVA

> tr\_A = c(rep(0,11),rep(1,10))

> tr\_B = c(rep(0,5),rep(1,6),rep(0,5),rep(1,5))

>

> summary(aov(avgs ~ tr\_A\*tr\_B))

Df Sum Sq Mean Sq F value Pr(>F)

tr\_A 1 0.1371 0.1371 6.823 0.018221 \*

tr\_B 1 0.3205 0.3205 15.954 0.000939 \*\*\*

tr\_A:tr\_B 1 0.0530 0.0530 2.636 0.122845

Residuals 17 0.3415 0.0201

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> summary(aov(avgs ~ tr\_A+tr\_B))

Df Sum Sq Mean Sq F value Pr(>F)

tr\_A 1 0.1371 0.1371 6.254 0.02227 \*

tr\_B 1 0.3205 0.3205 14.625 0.00124 \*\*

Residuals 18 0.3945 0.0219

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> #Assumptions of ANOVA

> ###(1)Normality

> #Anderson Darling test

> #install.packages('nortest')

> library('nortest')

> #H0 : y\_i/n\_i ~ Normal dist

> result\_AD\_test = ad.test(avgs)

> result\_AD\_test$p.value <= 0.05 #FALSE => do not reject H0

[1] FALSE

>

> ###(2)Homogeniety of Variance

> #H0 : y\_i/n\_i's have Homogeniety of variance(Seed)

> group\_vec1 = c(1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2)

> result\_BL\_test1 = bartlett.test(avgs~group\_vec1)

> result\_BL\_test1$p.value <= 0.05 #FALSE => do not reject H0

[1] FALSE

> #H0 : y\_i/n\_i's have Homogeniety of variance(Extract)

> group\_vec2 = c(1,1,1,1,1,2,2,2,2,2,2,1,1,1,1,1,2,2,2,2,2)

> result\_BL\_test2 = bartlett.test(avgs~group\_vec2)

> result\_BL\_test2$p.value <= 0.05 #FALSE => do not reject H0

[1] FALSE

>

> ###(3)Independence

>

>

> #(c) Establish an appropriate logistic regression model to the above data. Interpret the result.

> #Logistic Regression

> input = cbind(y\_i, n\_i-y\_i)

> Logistic\_model1 = glm(input~tr\_A\*tr\_B,family=binomial)

> summary(Logistic\_model1)

Call:

glm(formula = input ~ tr\_A \* tr\_B, family = binomial)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.01617 -1.24398 0.05995 0.84695 2.12123

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.5582 0.1260 -4.429 9.46e-06 \*\*\*

tr\_A 0.1459 0.2232 0.654 0.5132

tr\_B 1.3182 0.1775 7.428 1.10e-13 \*\*\*

tr\_A:tr\_B -0.7781 0.3064 -2.539 0.0111 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 98.719 on 20 degrees of freedom

Residual deviance: 33.278 on 17 degrees of freedom

AIC: 117.87

Number of Fisher Scoring iterations: 4

>

> Logistic\_model2 = glm(input~tr\_B,family=binomial)

> summary(Logistic\_model2)

Call:

glm(formula = input ~ tr\_B, family = binomial)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.70034 -1.26256 -0.08254 1.05562 2.38253

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.5122 0.1039 -4.927 8.34e-07 \*\*\*

tr\_B 1.0574 0.1438 7.353 1.93e-13 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 98.719 on 20 degrees of freedom

Residual deviance: 42.751 on 19 degrees of freedom

AIC: 123.35

Number of Fisher Scoring iterations: 3