Assignment 3

Last name: Du
First name: Min
Student ID: 1002602230

Part 1

1. Applied Chemistry.
2. Schnurr, P. J., Molenda, O., Edwards, E., Espie, G. S., & Allen, D. G. (2016). Improved biomass productivity in algal biofilms through synergistic interactions between photon flux density and carbon dioxide concentration. Bioresource Technology, 219, 72-79.
3. Department of Chemical Engineering.
4. https://doi.org/10.1016/j.biortech.2016.06.129
5. R (1.15.1), StatPlus.
6. From experiment.
7. Yes, the author presents a summary statistics of statistically significant coefficients that describe surface response model used to determine the impact of [CO2]*[PFD] on biomass productivity.
8. Yes, the authors use tests. For example, Biofilm biomass productivities at various CO2 concentrations were determined by linear regression analysis (95% confidence intervals) of plots of biofilm yields across the time course of the experiment.
9. The values were reported to 2 decimal places.
10. Linear regression analysis.

Part 2

Question 1

(a)

```
##
##          no yes
##   female 12  26
##   male    8  44
```

```
##
##  Pearson's Chi-squared test
##
## data:  sex_like2230
## X-squared = 3.3314, df = 1, p-value = 0.06797
```

H_0 = Sex and Student's preference for playing video games are independent.

H_a = Sex and Student's preference for playing video games are not independent.

First P-value = 0.0680, which is smaller than significance level 0.1, reject H_0, there is evidence of association between sex and student's preference for playing video games.

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_like2230
## p-value = 0.07824
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8195672 8.1070182
## sample estimates:
## odds ratio
##   2.511179
```

H_0 = Sex and Student's preference for playing video games are independent.

H_a = Sex and Student's preference for playing video games are not independent.

Second P-value = 0.0782, smaller than significance level 0.1, reject H_0, there is evidence of association between sex and student's preference for playing video games.

(b)

```
##
##          no yes
##   female  5   4
##   male    1  21
```

```
##
```

```
##  Pearson's Chi-squared test
##
## data:  gradeA_2230
## X-squared = 10.648, df = 1, p-value = 0.001102
```

H_0 = Sex and a student's preference for grade A are independent.

H_a = Sex and a student's preference for grade A are not independent.

When grade is A, p-value is 0.0011, which is smaller than 0.1,reject H_0, Sex and a student's preference for grade A are not independent at 0.1 significance level.

```
##
##          no yes
##   female  7  22
##   male    7  23
```

```
##
##  Pearson's Chi-squared test
##
## data:  gradenA_2230
## X-squared = 0.0052746, df = 1, p-value = 0.9421
```

H_0 = Sex and a student's preference for grade nA are independent.

H_a = Sex and a student's preference for grade nA are not independent.

When grade is nA, p-value is 0.9421, which is greater than 0.1, do not reject H_0, Sex and a student's preference for grade nA are independent at 0.1 significance level.

Therefore, the association between sex and students' preference for playing video games change with grade expected.

Question 2 (a)

model2.1: $log(\frac{\pi}{1-\pi})$=$\beta_0 + \beta_1 I_{male} + \beta_2 I_{gradenA} + \beta_3 I_{male} * I_{gradenA}$

model2.2: $log(\frac{\pi}{1-\pi})$=$\beta_0 + \beta_1 I_{male} + \beta_2 I_{gradenA}$

$I_{male}\{^{=1, Male}_{=0, Female}$

$I_{gradenA}\{^{=1, gradenA}_{=0, gradeA}$

$I_{male} * I_{gradenA}\{^{=1, male, gradenA}_{=0, Otherwise}$

Test1:

H_0: the coefficient of interaction term is 0. ($\beta_3 = 0$)

H_a: the coefficient of interaction term is not 0. ($\beta_3 \neq 0$)

```
## Analysis of Deviance Table
##
## Model 1: like ~ sex + grade + sex * grade
```

```
## Model 2: like ~ sex + grade
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        86     85.152
## 2        87     92.031 -1  -6.8788 0.008723 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-value=0.0087, smaller than 0.1, reject H_0, there is significant interaction effect, model 2.1 is better.

Test2:

H_0: the coefficient of interaction term is 0. ($\beta_3 = 0$)

H_a: the coefficient of interaction term is not 0. ($\beta_3 \neq 0$)

```
##
## Call:
## glm(formula = like ~ sex + grade + sex * grade, family = binomial(link = "logit"),
##     data = video2230)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4864   0.3050   0.7290   0.7433   1.2735
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.2231     0.6708  -0.333  0.73940
## sexmale            3.2677     1.2237   2.670  0.00758 **
## gradenA            1.3683     0.7989   1.713  0.08679 .
## sexmale:gradenA  -3.2232     1.3682  -2.356  0.01848 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 95.347  on 89  degrees of freedom
## Residual deviance: 85.152  on 86  degrees of freedom
## AIC: 93.152
##
## Number of Fisher Scoring iterations: 5
```

In summary, P-value of interaction term is $0.0185 < 0.1$, reject H_0, the interaction term is significant, model2.1 includes interaction term is better.

Two tests give the same result that there exists interaction effect. Therefore, model2.1 includes interaction term is better.

(b)

There is significant interaction effect between sex and grade. The probability of students' preference for playing video games between male and female differ with grade type.

In this context, Sex effect is affected by grade, which is consistent with part 1(b), so they are agree.

Question 3 (a)

```
##   like    sex grade count
## 1   no female     A     5
## 2  yes female     A     4
## 3   no   male     A     1
## 4  yes   male     A    21
## 5   no female    nA     7
## 6  yes female    nA    22
## 7   no   male    nA     7
## 8  yes   male    nA    23
```

model3.1: $\log(\mu_{ijk}) \sim \beta_0 + \beta_1 I_{[sex=male]} + \beta_2 I_{[grade=nA]} + \beta_3 I_{[like=yes]} + \beta_4 I_{[sex=male]} * I_{[grade=nA]} + \beta_5 I_{[sex=male]} * I_{[like=yes]} + \beta_6 I_{[grade=nA]} * I_{[like=yes]} + \beta_7 I_{[sex=male]} * I_{[grade=nA]} * I_{[like=yes]}$

model3.2: $\log(\mu_{ijk}) \sim \beta_0 + \beta_1 I_{[sex=male]} + \beta_2 I_{[grade=nA]} + \beta_3 I_{[like=yes]} + \beta_4 I_{[sex=male]} * I_{[grade=nA]} + \beta_5 I_{[sex=male]} * I_{[like=yes]} + \beta_6 I_{[grade=nA]} * I_{[like=yes]}$

$I_{[sex=male]}\{^{=1,Male}_{=0,Female}$

$I_{[grade=gradenA]}\{^{=1,gradenA}_{=0,gradeA}$

$I_{[like=yes]}\{^{=1,like=yes}_{=0,like=no}$

$I_{male} * I_{gradenA}\{^{=1,male,gradenA}_{=0,Otherwise}$

$I_{male} * I_{yes}\{^{=1,male,yes}_{=0,Otherwise}$

$I_{gradenA} * I_{yes}\{^{=1,gradenA,yes}_{=0,Otherwise}$

$I_{[sex=male]} * I_{[grade=nA]} * I_{[like=yes]}\{^{=1,male,gradenA,yes}_{=0,Otherwise}$

(b)

i.

Model	Deviance
model31	-5.773162e-15
model32	6.878764
model21	85.15215
model22	92.03091

Deviance for poisson models are much lower.

ii.

Model	Test Statistics	Distribution
Logistic Regression	6.8788-(-5.7732e-15)=6.87	$X^2(1)$
Poisson Regression	92.031-85.152=6.87	$X^2(1)$

Note that the distribution for using wald test in part2 and and part3 is the same, follows $X^2(1)$.

The Test Statistics in part2 under logistic regression is 6.8788-(-5.7732e-15)=6.87.

The Test Statistics in part3 under poisson distribution is 92.031-85.152=6.87.

Two test statistics are the same.

According to Wald test,$H_0 : \beta_7 = 0$, $H_a : \beta_7 \neq 0$, p-value is 0.018, smaller than significance level 0.1, reject H_0, $\beta_7 \neq 0$.

Under the same distribition and the same test statistics. The wald test gives the same result.

iii.

In poisson regression, increase x_{ijk} by one unit, holding other predictors constant.μ_j changes by a factor of $e^{\beta_{ijk}}$, poisson regression focus on investigating the count affected by predictors.

In logistic model, increase x_{ijk} by one unit, holding other predictors constant.μ_j changes log odd ratio by β, logistic model focus on investigating the probability of like affected by predictors.

Appendix

(Part 2)

1.

(a)

```
video2230=read.csv("/Users/mindu/Desktop/STA303/Assignment3/video.csv")
attach(video2230)
sex_like2230=table(video2230$sex,video2230$like)
sex_like2230
```

```
chisq.test(sex_like2230,correct = FALSE)
```

```
fisher.test(sex_like2230)
```

(b)

```
gradeA_2230= table(video2230$sex[video2230$grade == "A"], video2230$like[video2230$grade
gradeA_2230
chisq.test(gradeA_2230,correct = F)
gradenA_2230 = table(video2230$sex[video2230$grade == "nA"], video2230$like[video2230$gr
gradenA_2230
chisq.test(gradenA_2230, correct = F)
```

2.

(a) Test1:

```
model21_2230 =glm(like~sex+grade+sex * grade, family = binomial(link = "logit") ,data =
model22_2230 =glm(like~sex+grade, family = binomial(link = "logit") ,data = video2230)
anova(model21_2230,model22_2230,test="Chisq")
```

Test2:

```
summary(model21_2230)
```

3.

(a)

```
count2230=table(video2230)
count_2230=as.data.frame(count2230)
colnames(count_2230)[4]= "count"
count_2230
```

(b)

i.

```r
model31_2230 = glm(count ~  sex + grade + like + sex*grade + sex*like + grade*like + sex
model32_2230 = glm(count ~  sex + grade + like + sex*grade + sex*like + grade*like, fami
deviance(model31_2230)
deviance(model32_2230)
deviance(model21_2230)
deviance(model22_2230)
```

ii.

```r
summary(model31_2230)
summary(model32_2230)
summary(model21_2230)
summary(model22_2230)
library(aod)
wald.test(Sigma = vcov(model31_2230),b=coef(model31_2230),Term = 8)
```