Oscar Huang
April 05, 2017

Statistics 516
Homework 03

Student: Oscar Huang

Date: 2017/04/05

Oscar Huang
April 05, 2017

Impact on Pesticides on Skylark Reproductivity

1.

a. Code and output:

```
options(digits=4)
#building saturated model
skylark$batch<-factor(1:nrow(skylark))
m.sat <- glm(count ~ batch, family = "poisson", data = skylark)

m.pos<- glm(count ~ spray +field + year ,family = "poisson", data = skylark)
summary(m.pos)
```

```
Call:
glm(formula = count ~ spray + field + year, family = "poisson",
    data = skylark)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.8682  -0.7351   0.0244   0.6697   1.8797

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.43094    0.13262   25.87  < 2e-16 ***
sprayyes    -0.45613    0.09385   -4.86  1.2e-06 ***
fieldKr      0.04909    0.12672    0.39   0.6985
fieldKu      0.00496    0.12800    0.04   0.9691
fieldRd     -0.17905    0.13417   -1.33   0.1820
year1993     0.46262    0.13064    3.54   0.0004 ***
year1994     0.06002    0.14149    0.42   0.6714
year1995     0.32728    0.13411    2.44   0.0147 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 69.960  on 15  degrees of freedom
Residual deviance: 18.984  on  8  degrees of freedom
AIC: 118.3

Number of Fisher Scoring iterations: 4
```

```
anova(m.pos,m.sat,test = "LRT")
```

```
Analysis of Deviance Table

Model 1: count ~ spray + field + year
Model 2: count ~ batch
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         8        19
2         0         0  8       19    0.015 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Oscar Huang
April 05, 2017

```
plot(predict(m.pos),rstudent(m.pos))
```

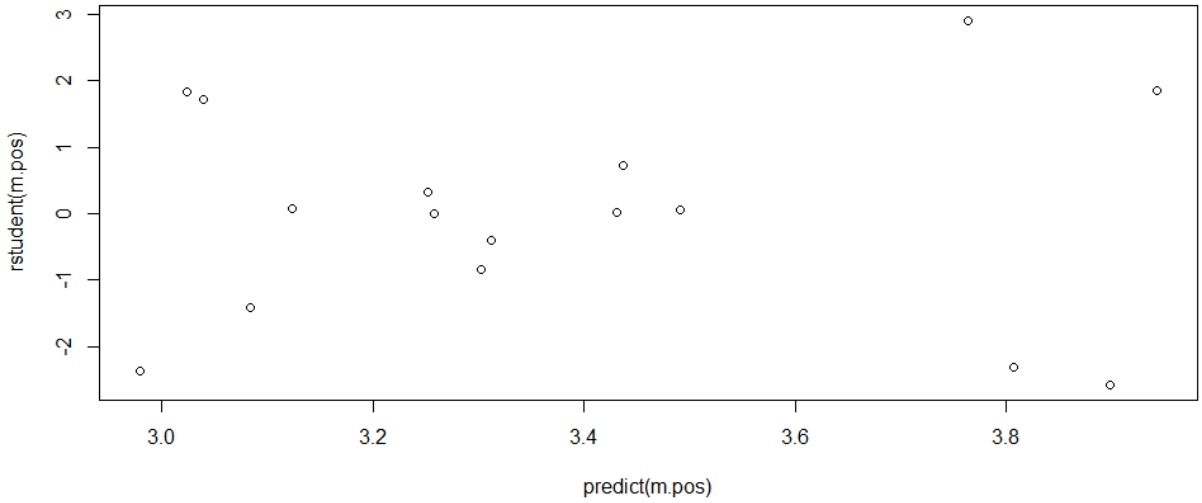

Figure 1 The residuals with different prediction value of the poisson model.

```
m.quasipos <- glm(count ~ spray + field + year, family = quasipoisson, data =
skylark)
plot(predict(m.quasipos),rstudent(m.quasipos))
```

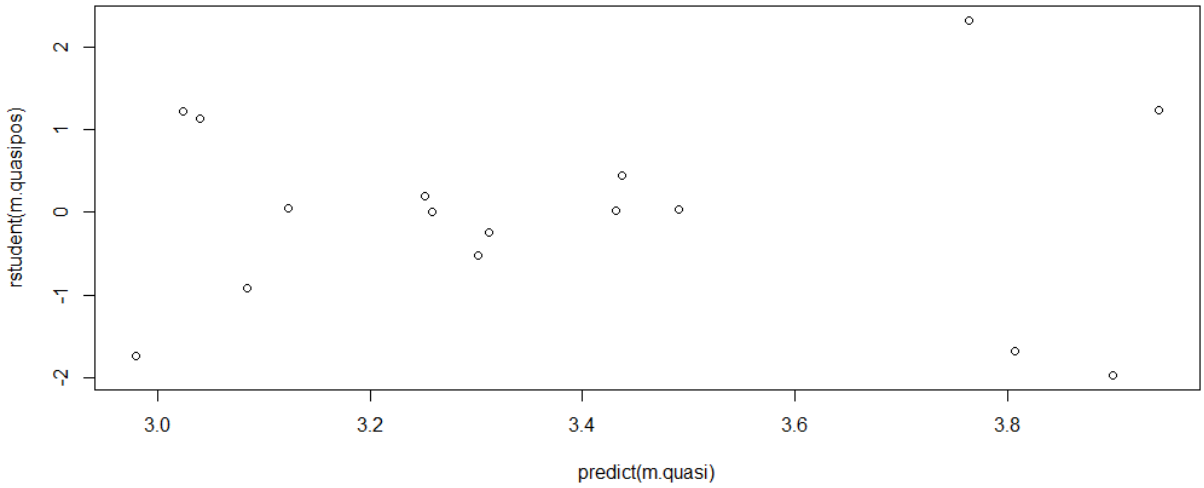

Figure 2 The residuals with different prediction value of the quasi-poisson model.

```
summary(m.quasi)$coefficients
```

```
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  3.430943     0.2029 16.91025 1.517e-07
sprayyes    -0.456126     0.1436 -3.17687 1.306e-02
fieldKr      0.049089     0.1939  0.25322 8.065e-01
fieldKu      0.004964     0.1958  0.02535 9.804e-01
fieldRd     -0.179048     0.2053 -0.87233 4.084e-01
year1993     0.462623     0.1999  2.31467 4.933e-02
year1994     0.060018     0.2164  0.27728 7.886e-01
year1995     0.327281     0.2052  1.59520 1.493e-01
```

Oscar Huang
April 05, 2017

b. Discussion:

When using the deviance of the residuals to test the Poisson regression model, the p-value was 0.015, which was lower than the α value we usually use (0.05). This means the deviance is not likely to happen under the null hypothesis ($D=2\left(\log L_s - logL\right)=0$) is true. Where L and L_s are the likelihoods of the Poisson regression model and a saturated model. When plotting the Figure 1, the residuals with different prediction value of the Poisson model, there are some residuals higher than 2 or lower than -2. This also shows the evident of overdispersion

When I apply the quasi-Poisson regression model and plot the Figure 2, The residuals with different prediction value of the quasi-Poisson model, almost all the residuals are within -2 and 2, which means the model has no overdispersion.

$$E\left(Y_i\right)=\exp\left(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}+\beta_5 x_{i5}+\beta_6 x_{i6}+\beta_7 x_{i7}\right)$$

$x_{i1}=1 if\ the\ pesticide\ is\ applied\ at\ i-th\ sample\ ;\ x_{i1}=0\ other\ wise$
$x_{i2}=1 if\ the\ i-th\ sample\ is\in field\ \mathrm{Kr}\ ;\ x_{i2}=0\ other\ wise$
$x_{i3}=1 if\ the\ i-th\ sample\ is\in field\ \mathrm{Ku}\ ;\ x_{i3}=0\ other\ wise$
$x_{i4}=1 if\ the\ i-th\ sample\ is\in field\ \mathrm{Rd}\ ;\ x_{i4}=0\ other\ wise$
$x_{i5}=1 if\ the\ i-th\ sample\ is\ at\ year\ 1993\ ;\ x_{i5}=0\ other\ wise$
$x_{i6}=1 if\ the\ i-th\ sample\ is\ at\ year\ 1994\ ;\ x_{i6}=0\ other\ wise$
$x_{i7}=1 if\ the\ i-th\ sample\ is\ at\ year\ 1995\ ;\ x_{i7}=0\ other\ wise$

If the pesticide was not applied, the expected number of the skylark fledglings will be $\exp\left(\beta_0+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}+\beta_5 x_{i5}+\beta_6 x_{i6}+\beta_7 x_{i7}\right)$, and if the pesticide was applied, the expected number of the skylark fledglings will be $\exp\left(\beta_0-0.456126+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}+\beta_5 x_{i5}+\beta_6 x_{i6}+\beta_7 x_{i7}\right)$. Which means when the pesticide was applied, the expected number of the skylark fledglings will decrease by a factor of exp(-0.456126)=0.6337, which also means the number will decrease 36.63%.

Oscar Huang
April 05, 2017

Moth Coloration and Natural Selection

1.

a. Code and output:

```
m<-glm(cbind(Removed,Placed-Removed) ~ Morph * Distance, family = binomial,
data=case2102)
contrast(m,
        a = list(Morph = c("dark","light"), Distance = 50),
        b = list(Morph = c("dark","light"), Distance = 0),
        cnames = c("Dark","Light"), tf = exp)
```

```
      estimate     se  lower upper tvalue  df   pvalue
Dark    2.5222 0.2823 1.4505 4.386  3.277 Inf 0.001048
Light   0.6286 0.2894 0.3564 1.108 -1.604 Inf 0.108629
```

```
#Plot
mydata_dark <- data.frame(Distance = seq(0, 50, by = 0.5),Morph = "dark")
mydata_dark$logo<- predict(m, newdata = mydata_dark)
mydata_dark$odds <- exp(mydata_dark$logo)
mydata_light <- data.frame(Distance = seq(0, 50, by = 0.5),Morph = "light")
mydata_light$logo<- predict(m, newdata = mydata_light)
mydata_light$odds <- exp(mydata_light$logo)
p<-ggplot(case2102, aes(x=Distance,y=Removed/(Placed-Removed)))
p<-p+geom_point(aes(fill=Morph),shape=21)
p<-p+scale_fill_manual(values=c("black","white"))
p<-p+ylab("Odds of Removed Moths")
p<-p+xlab("Distance from Liverpool (km)")
p<-p + geom_line(aes(y = odds), data = mydata_dark)
p<-p + geom_line(aes(y = odds), data = mydata_light,linetype=2)
plot(p)
```

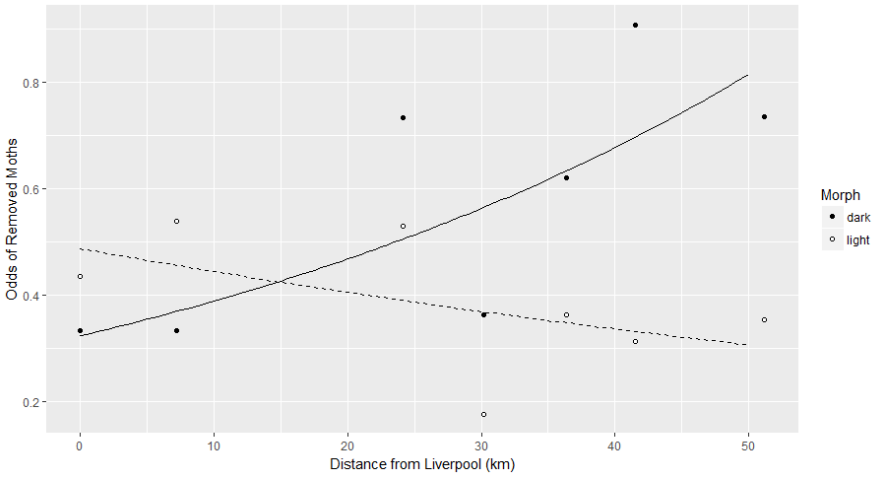

Figure 3 The odds of moths being removed with the different distance from Liverpool. The solid line represents the prediction curve of the dark moths, and the dash line indicates the prediction curve of the light moths.

Oscar Huang
April 05, 2017

b. Discussion:

Odds ratio for dark moth: 2.5222; Odds ratio for light moth: 0.6286

When the distance increased by 50 km, the odds that a dark moth will be removed changed by a factor of 2.5222, and the odds that a light moth will be removed changed by a factor of 0.6286.

These numbers agree with the Figure 3. The odds of dark moths being removed at 0 and 50 km are about 0.32 and 0.81, and the odds ratio is 0.81/0.32 =2.53 ≈ 2.5222. The odds of light moths being removed at 0 and 50 km are about 0.48 and 0.30, and the odds ratio is 0.30/0.48 =0.625 ≈ 0.6286.

2.

a. Code and output:

#margeff 50-0

```
margeff(m,
        a = list(Morph = c("dark","light"), Distance = 50),
        b = list(Morph = c("dark","light"), Distance = 0),
        cnames = c("Dark","Light"))
```

```
      estimate      se    lower   upper tvalue  df    pvalue
Dark   0.20486 0.05927  0.08869 0.32103  3.456 Inf 0.0005479
Light -0.09321 0.05867 -0.20820 0.02177 -1.589 Inf 0.1120949
```

#margeff 50-0 percent

```
margeff(m,
        a = list(Morph = c("dark","light"), Distance = 50),
        b = list(Morph = c("dark","light"), Distance = 0),
        cnames = c("Dark","Light"),type="percent")
```

```
      estimate    se  lower    upper tvalue  df  pvalue
Dark     83.84 35.29  14.68 153.0026  2.376 Inf 0.01751
Light   -28.43 14.80 -57.43   0.5781 -1.921 Inf 0.05474
```

#margeff at 25

```
margeff(m, delta = 0.001,
        a = list(Distance = 25 + 0.001, Morph = c("dark","light")),
        b = list(Distance = 25, Morph = c("dark","light")),
        cnames = c("Dark","Light"))
```

```
       estimate       se     lower     upper tvalue  df    pvalue
Dark   0.004148 0.001229  0.001738 0.0065575  3.374 Inf 0.0007421
Light -0.001868 0.001180 -0.004181 0.0004452 -1.583 Inf 0.1134827
```

#margeff at dark - light at 0 and 50

```
margeff(m,
        a = list(Morph = "dark", Distance = c(0,50)),
        b = list(Morph = "light", Distance = c(0,50)),
```

```
                cnames = c("0","50"))
```

```
  estimate      se   lower   upper tvalue  df    pvalue
0  -0.08354 0.05561 -0.1925 0.02545 -1.502 Inf 1.330e-01
50  0.21453 0.04654  0.1233 0.30575  4.609 Inf 4.038e-06
```

#plot

```
mydata_dark <- data.frame(Distance = seq(0, 50, by = 0.1),Morph = "dark")
mydata_light <- data.frame(Distance = seq(0, 50, by = 0.1),Morph = "light")
mydata_dark$prob <- predict(m, newdata = mydata_dark,type="response")
mydata_light$prob <- predict(m, newdata = mydata_light,type = "response")

p<-ggplot(case2102, aes(x=Distance,y=Removed/Placed))
p<-p+geom_point(aes(fill=Morph),shape=21)
p<-p+scale_fill_manual(values=c("black","white"))
p<-p+ylab("Probability of Removed Moths")
p<-p+xlab("Distance from Liverpool (km)")
p<-p + geom_line(aes(y = prob), data = mydata_dark)
p<-p + geom_line(aes(y = prob), data = mydata_light, linetype = 2)
plot(p)
```

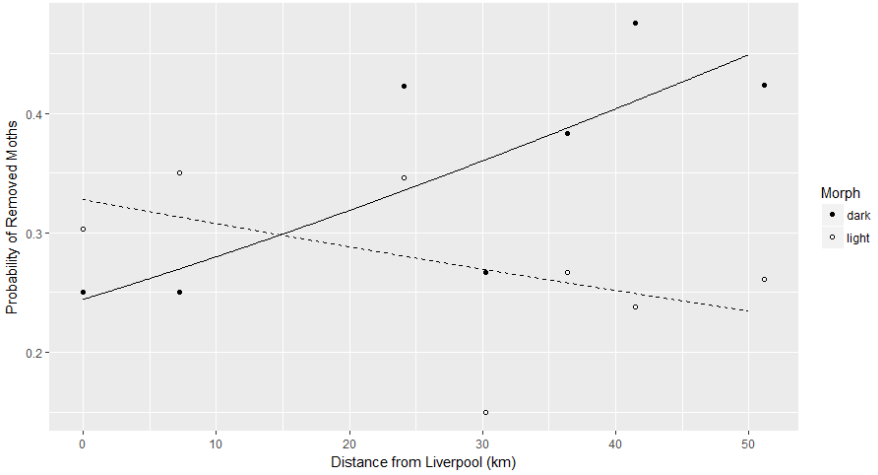

Figure 4 The probability of moths being removed with the different distance from Liverpool. The solid line represents the prediction curve of the dark moths, and the dash line indicates the prediction curve of the light moths.

b. Discussion:

The discrete marginal effects of increasing distance from 0 to 50 km on the probability of dark and light moth removal are 0.2048 and -0.0932, which means when the distance increased from 0 to 50 km, the probability of the dark and light moth being removed will increased 0.2048 (increase 83.84%) and decreased 0.0932 (decrease 28.43%).

The estimated instantaneous rate of change in the probability of dark and light moth removal at 25 km are 0.004148 and -0.001868, which means at 25 km the slope

of the prediction curve of dark and light moths being removed are 0.4148% per km and 0.1868% per km.

The discrete marginal effects for the difference in the probability of moth removal at 0 and 50 km are -0.08354 and 0.21453 (dark moth – light moth). Which means the dark moth has 8.354% lower probability to be removed at 0km, and 21.45% higher probability to be removed at 50 km. All the numbers obtained agree with Figure 3.

Oscar Huang
April 05, 2017

Cancer Death Rate of Atomic Bomb Survivors

1.

a. Code and output:

```
data$batch<-factor(1:nrow(data))
m.sat <- glm(Deaths ~ batch, family = "poisson", data = data)
m<-glm(Deaths ~ YearsAfter+Exposure,offset = log(AtRisk), family = poisson, data
= data)
plot(predict(m),rstudent(m))
```

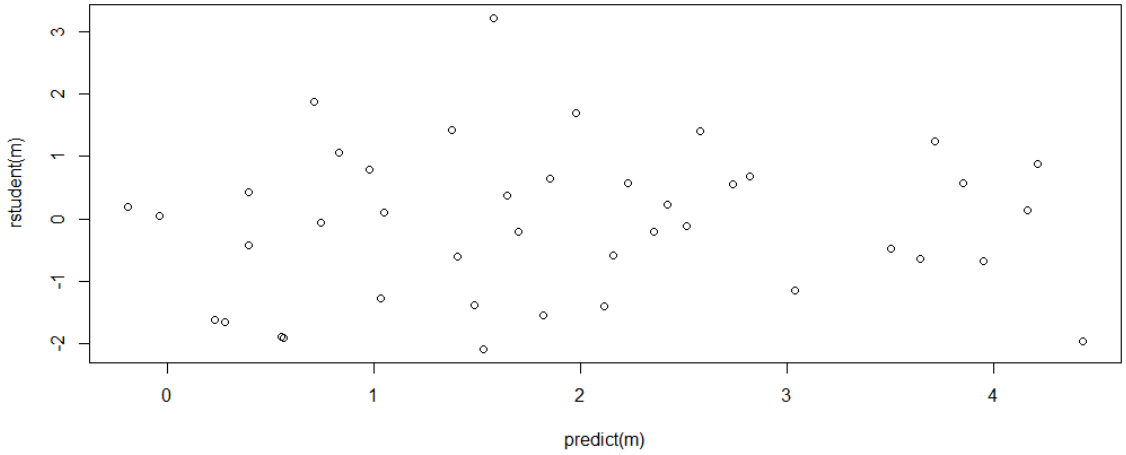

Figure 5 The residuals with different prediction value of the poisson model.

```
anova(m,m.sat,test = "LRT")
```

```
Model 1: Deaths ~ YearsAfter + Exposure
Model 2: Deaths ~ batch
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        34       50.1
2         0        0.0 34     50.1    0.037 *
```

```
m.quasi<-glm(Deaths ~ YearsAfter+Exposure, offset=log(AtRisk), family =
quasipoisson, data = data)
plot(predict(m.quasi),rstudent(m.quasi))
```

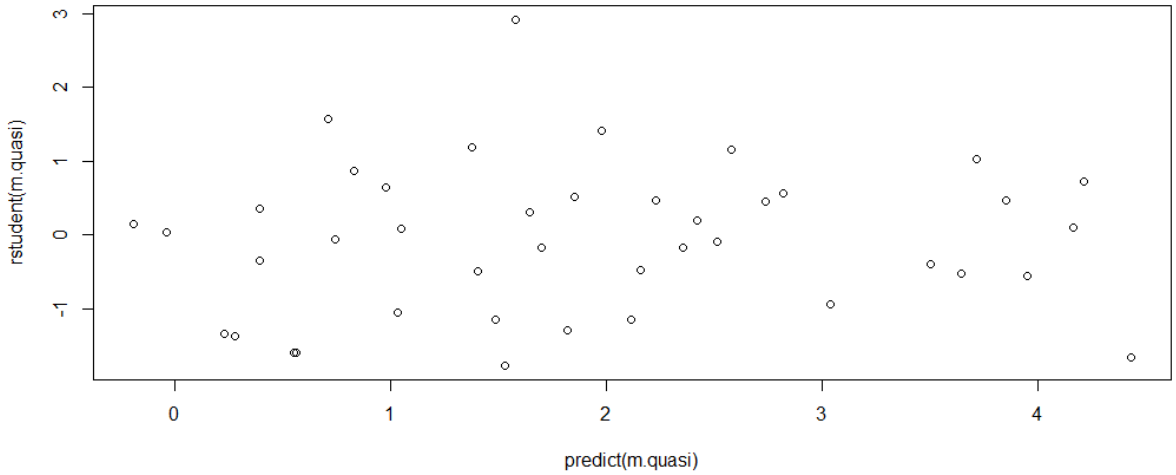

Figure 6 The residuals with different prediction value of the quasi-poisson model.

Oscar Huang
April 05, 2017

Oscar Huang
April 05, 2017

```
mydata <- expand.grid(YearsAfter =
c("0to7","8to11","12to15","16to19","20to23","24to27","28to31"),
                      Exposure = seq(0,400, by=1), AtRisk = 1)
mydata$y <- predict(m.quasi, newdata = mydata,
                        type = "response")

p<-ggplot(data,aes(x=Exposure, y=Deaths/AtRisk))
p<-p+geom_line()+facet_wrap(~YearsAfter, ncol=7)
p<-p+geom_point(aes(size = AtRisk), shape=21, fill = "white")
p<- p+xlim(-25,400)
p<- p+ labs(size = "Person-Years", y = "Cancer Death Rate (Deaths per Person-
Year)",x="Radiation Dose (rads)")
p<- p+ scale_size(breaks=c(1000,5000,10000,15000,20000))
p<-p+theme(axis.text.x = element_text(angle=45,vjust=1,hjust=1))
p<-p + geom_line(aes(y =y), data = mydata, linetype =2)
plot(p)
```

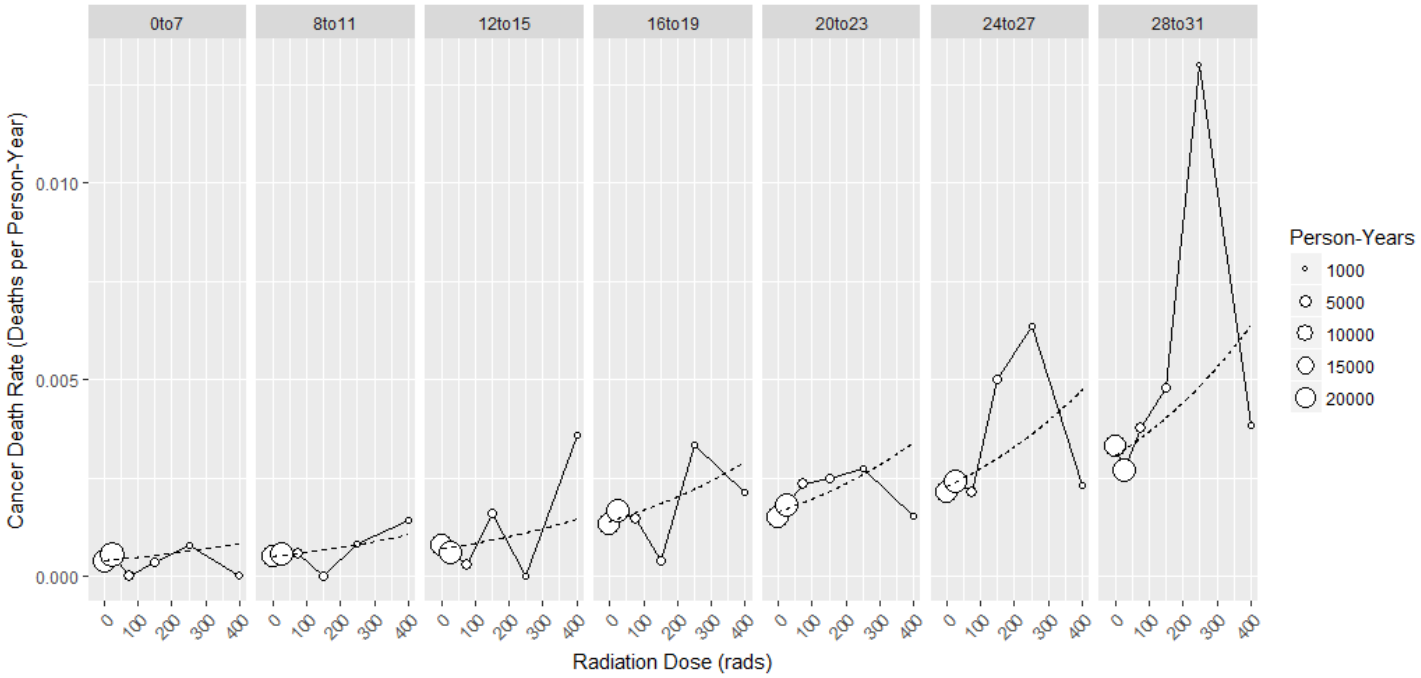

Figure 7 The cancer death rate with the different radiation dose and years after exposure. The dash line indicates the prediction curve of the quasi-poisson model.

```
contrast(m.quasi,
                a = list(YearsAfter = "0to7", Exposure = 100, AtRisk = 1),
                b = list(YearsAfter = "0to7", Exposure = 0, AtRisk = 1), tf = exp)
```

```
estimate      se lower upper tvalue df   pvalue
   1.201 0.05236  1.08 1.336  3.498 34 0.001328
```

Oscar Huang
April 05, 2017

```
exp(summary(m.quasi)$coefficients)
```

```
              Estimate Std. Error   t value Pr(>|t|)
(Intercept)     0.0004017      1.250 5.726e-16    1.000
YearsAfter8to11  1.2627945      1.352 2.169e+00    1.559
YearsAfter12to15 1.7361996      1.327 7.040e+00    1.061
YearsAfter16to19 3.4842188      1.289 1.356e+02    1.000
YearsAfter20to23 4.0709552      1.284 2.724e+02    1.000
YearsAfter24to27 5.6783267      1.275 1.272e+03    1.000
YearsAfter28to31 7.6228003      1.269 5.066e+03    1.000
Exposure        1.0018333     1.001 3.305e+01    1.001
```

```
1.0018333^100
```

```
[1] 1.201
```

b. Discussion:

The p-value of the deviance of residuals test on the Poisson regression model was 0.37, which was close to α value we usually use (0.05), and the Figure 5 also shows most of the residuals are in the range of -2 and 2, which means the model has a little bit over dispersion. When applying the quasi-Poisson regression model, Figure 6 shows the range of the residuals decreased (no significant different from the Poisson model).

Since the year after exposing does not interact with the dose, I only used the data set with 0 to 7 years after exposure in the contrast function to predict the dose effects on the death rate. The estimated value from the contrast function was 1.201, which means when the dose increase 100 rad, the death rate will increase by a factor of 1.201 (increased 20.1%). We can also calculate this number from the exp(summary(m.quasi)$coefficients) function. The estimated increasing of death rate while the exposure increased 1 rad was 1.0018333. the effect of increasing 100 rad will be $1.0018333^{100} = 1.201$, which agrees with the data from the contrast function.

Oscar Huang
April 05, 2017

Producing Embryogenic Anthers

1.

a. Code and output:

```
m2 <- glm(y/n ~ force * storage,data = anthers, family = tweedie(link.power = -3,
var.power = 2))
summary(m2)
```

```
Call:
glm(formula = y/n ~ force * storage, family = tweedie(link.power = -3,
    var.power = 2), data = anthers)

Deviance Residuals:
      1       2       3       4       5       6
 0.00757 -0.01205  0.00440  0.01179 -0.02658  0.01444

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.30025    0.23309    9.87    0.010 *
force               0.01083    0.00157    6.90    0.020 *
storagecontrol      4.17706    0.54622    7.65    0.017 *
force:storagecontrol -0.00964    0.00278   -3.47    0.074 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Tweedie family taken to be 0.000633)

    Null deviance: 0.0818936  on 5  degrees of freedom
Residual deviance: 0.0012762  on 2  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 3
```

```
plot(predict(m2),rstudent(m2))
```

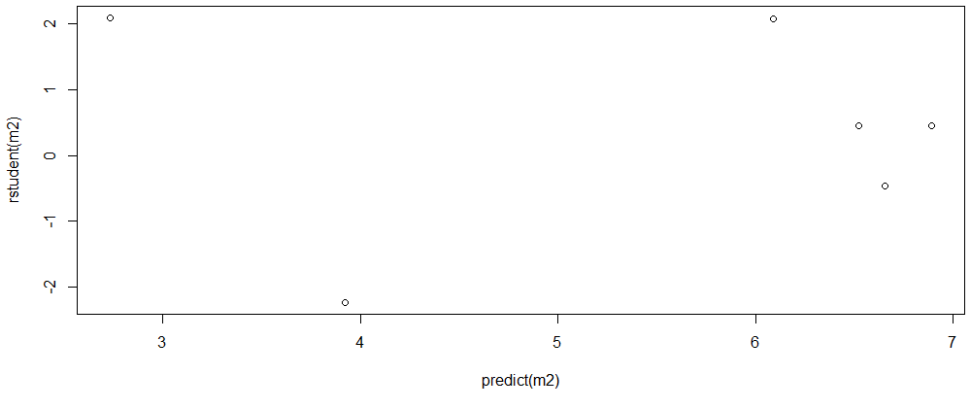

Figure 8 The residuals with different prediction value of the tweedie model.

```
d<- expand.grid(force = seq(40,350, length = 100), storage = c("cold",
"control"),n=1)
```

Oscar Huang
April 05, 2017

```
d$yhat<-predict(m2,newdata = d, type = "response")

p<-ggplot(anthers, aes(x=force,y=y/n))
p<-p+geom_line(aes(y=yhat, group=storage), data =d)
p<-p+geom_point(aes(size = n, fill = storage),  shape = 21)
p<-p+scale_fill_manual(values = c("white",grey(0.75)))
p<-p+scale_size(breaks=c(50,150,250,350))
p<-p+labs(y="Expected Proportion of \n Embryogenic Anthers", x="Force (g)", size
= "Number of \nAnthers", fill = "Storage \n Condition")
plot(p)
```

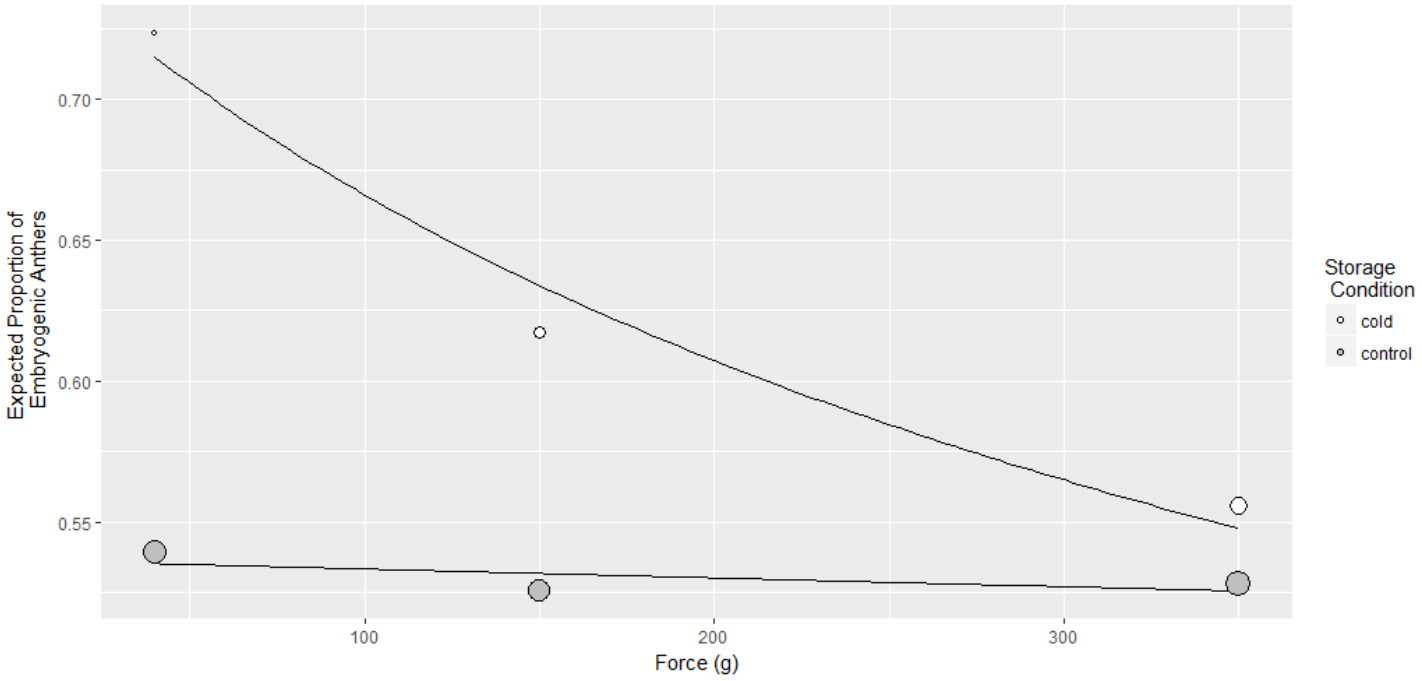

Figure 9 The expected proportion of embryogenic anthers with different storage condition, g force.

b. Discussion:

The tweedie model was used to fit the data, and Figure 8 shows there was no overdispersion in the model. The mean structure of the tweedie model is: $g\left[E\left(Y_i\right)\right]=\eta_i=E\left(Y_i\right)^{-3}$, hence $\eta_i^{-1/3}=E\left(Y_i\right)$.

$$E\left(Y_i\right)=\left(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2}\right)^{-1/3}$$

$x_{i1}=g\,force$

$x_{i2}=1\,if\ the\ storage\ condition\ of\ the\ i-th\ fruitfly\ was\in\ con\quad trol\quad group\,;\,x_{2i}=0\,other\ wise$

$$E\left(Y_i\right)=\left(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2}\right)^{\frac{-1}{3}}$$

$if\ the\ storage\ condition\ of\ the\ i-th\ fruitfly\ was\in\ con\quad trol\quad group$

Oscar Huang
April 05, 2017

$$E(Y_i)=(\beta_0+\beta_1 x_{i1})^{\frac{-1}{3}}$$
$$if\ the\ storage\ condition\ of\ the\ i-th\ fruitfly\ was \in\ cold\quad group$$

The effect of force on the expected proportion for “control” group at force = f was:

$$\frac{d}{df}(2.300+0.011*f+4.177+(-0.009)*f)^{\frac{-1}{3}}=\frac{-0.02(0.002f+6.477)^{\frac{-4}{3}}}{3}$$

The effect of force on the expected proportion for “cold” group at force = f was:

$$\frac{d}{df}(2.30+0.011*f)^{\frac{-1}{3}}=\frac{-0.011(0.011f+2.3)^{\frac{-4}{3}}}{3}$$

The storage condition effects on the sample at 40g was:

$$\frac{E_{cold}(Y_i)}{E_{control}(Y_i)}=\left(\frac{2.300+0.011*40}{2.300+0.011*40+4.177+(-0.009)*40}\right)^{\frac{-1}{3}}=1.3376$$

When stored the sample at cold condition with 40 *g*, the expected proportion of embryogenic anthers will increase by a factor of 1.3376 (33.76% increasing)

The storage condition effects on the sample at 150g was:

$$\frac{E_{cold}(Y_i)}{E_{control}(Y_i)}=\left(\frac{2.300+0.011*150}{2.300+0.011*150+4.177+(-0.009)*150}\right)^{\frac{-1}{3}}=1.1971$$

When stored the sample at cold condition with 150 *g*, the expected proportion of embryogenic anthers will increase by a factor of 1.1971 (19.71% increasing)

The storage condition effects on the sample at 350g was:

$$\frac{E_{cold}(Y_i)}{E_{control}(Y_i)}=\left(\frac{2.300+0.011*350}{2.300+0.011*350+4.177+(-0.009)*350}\right)^{\frac{-1}{3}}=1.0528$$

When stored the sample at cold condition with 350 *g*, the expected proportion of embryogenic anthers will increase by a factor of 1.0528 (5.28% increasing)