

# HW5\_\_answer

Guojing Wu

3/7/2019

## Problem 1

### a) Fit a poisson model (M1)

```
data_crab <- read.csv("HW5-crab.txt", sep = "")
fit.m1 <- glm(Sa ~ W, data = data_crab, family = poisson(link = log))
res.m1 <- summary(fit.m1)
```

The null model deviance  $D_{null}$  is 632.792 while the M1 model deviance  $D_{M1}$  is 567.879,  $D_{null} - D_{M1} \sim \chi_1^2$ . Hence we got the pvalue = 0, we reject the null hypothesis and state that the M1 model doesn't fit the data well.

We interpret that: the log rate ratio of number of satellites is 0.164 given one unit change in carapace width.

### b) Fit a model (M2) with W and Wt, compare it with M1

```
fit.m2 <- glm(Sa ~ W + Wt, data = data_crab, family = poisson(link = log))
res.m2 <- summary(fit.m2)
```

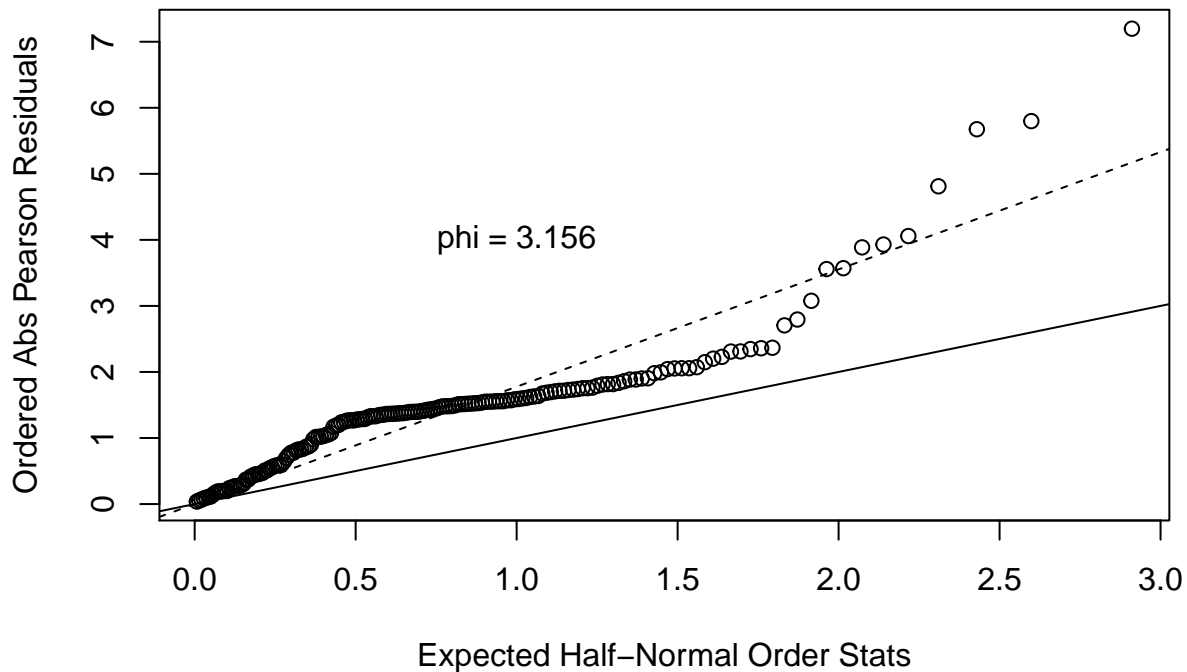
The M2 model deviance  $D_{M2}$  is 559.885,  $D_{M1} - D_{M2} \sim \chi_1^2$ . Hence we got the pvalue = 0.005, we failed to reject the null hypothesis and state that the M2 model fits the data better.

We interpret that:

- the log rate ratio of number of satellites is 0.046 given one unit change in carapace width.
- the log rate ratio of number of satellites is 0.447 given one unit change in weight.

### c) check over dispersion in M2

```
residual.m2 = residuals(fit.m2, type = 'pearson', data = data_crab)
G = sum(residual.m2 ^ 2) # calc dispersion param based on full model
phi = G / (173 - 3)
plot(qnorm((173 + 1:173 + 0.5) / (2 * 173 + 1.125)), sort(abs(residual.m2)), xlab = 'Expected Half-Normal',
     abline(a = 0, b = 1)
     abline(a = 0, b = sqrt(phi), lty = 2)
     text(1, 4, labels = paste("phi = ", round(phi, 3), sep = ""))
```



The half normal plot shows that M2 does exist over dispersion.

```
res.m2.dis <- summary(fit.m2, dispersion = phi)
```

We interpret that:

- the log rate ratio of number of satellites is 0.046 given one unit change in carapace width.
- the log rate ratio of number of satellites is 0.447 given one unit change in weight.

## Problem 2

a)

```
data_para <- read.csv("HW5-parasite.txt", sep = "") %>%
  mutate(Year = as.factor(Year),
         Area = as.factor(Area))
fit_para <- glm(Intensity ~ Area + Year + Length, data = data_para, family = poisson(link = log))
res_para <- summary(fit_para)
```

We interpret that:

- the log rate ratio of intensity is -0.212 given area change from 1 to 2.
- the log rate ratio of intensity is -0.117 given area change from 1 to 3.
- the log rate ratio of intensity is 1.405 given area change from 1 to 4.
- the log rate ratio of intensity is 0.67 given year change from 1999 to 2000.
- the log rate ratio of intensity is -0.218 given year change from 1999 to 2001.
- the log rate ratio of intensity is -0.028 for 1 unit change in length.

**b) test for goodness of fit in (a)**

```
pval = 1 - pchisq(res.para$deviance, 1191 - 7)
```

The value we got is 0, so we reject the null hypothesis and state that this model doesn't fit the data well.

**c) use zero-inflated model**

We assume that the area may be related with whether susceptible and build the model like this:

```
fit.zip <- zeroinfl(Intensity ~ Area + Year + Length | Area, data = data_para)
res.zip <- summary(fit.zip)
```

We interpret for count part:

- within the susceptible fish, the log rate ratio of intensity is 0.269 given area change from 1 to 2.
- within the susceptible fish, the log rate ratio of intensity is 0.147 given area change from 1 to 3.
- within the susceptible fish, the log rate ratio of intensity is 0.945 given area change from 1 to 4.
- within the susceptible fish, the log rate ratio of intensity is 0.392 given year change from 1999 to 2000.
- within the susceptible fish, the log rate ratio of intensity is -0.046 given year change from 1999 to 2001.
- within the susceptible fish, the log rate ratio of intensity is -0.037 for 1 unit change in length.

We interpret for binomial part:

- the log odds ratio of susceptible or not is 0.757 given area change from 1 to 2.
- the log odds ratio of susceptible or not is 0.686 given area change from 1 to 3.
- the log odds ratio of susceptible or not is -0.85 given area change from 1 to 4.