# Homework 3

Xinyi Lin 2/20/2019

### Problem 1

#### Question 1

We assume  $Y_1$  is the number of people get esophageal cancer given their daily alcohol consumption is equal to or larger than 80g,  $Y_0$  is the number of people get esophageal cancer given their daily alcohol consumption is less than 80g.

$$Y_0 \sim Bin(m_0, \pi_0)$$
  
 $Y_1 \sim Bin(m_1, \pi_1)$ 

Logistic model:

$$logit(\pi_i) = log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

where  $X_1$  represents daily alcohol consumption and  $X_2$  represents age.

By using R, we get following results

```
## (Intercept) -5.02344878 0.418223504 -12.011398 3.095687e-33
## prosp_data$expo 1.77999951 0.187086038 9.514336 1.828762e-21
## prosp_data$age 0.06157867 0.007290651 8.446251 3.008053e-17
So the fitted model:
```

 $logit(\pi_i) = log \frac{\pi_i}{1 - \pi_i} = -5.023 + 1.780X_1 + 0.062X_2$ 

 $\beta_0$  cannot be interpreted,  $\beta_1 = 1.780$  means the log odd ratio of esophageal cancer between people whoes daily alcohol consumption is equal to or lager than 80g and people whoes daily alcohol consumption is less than 80g given their ages are same is 1.780 and  $\beta_2 = 0.062$  means the log odd ration of esophageal cancer is 0.062 when age increases 1 and daily alcohol consumption keeps constant.

#### Problem 2

#### Question 1

Logistic model:

$$logit(\pi_i) = log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

where  $X_1 = 0$  represents O.aegyptiaca 73,  $X_1 = 1$  represents O.aegyptiaca 75 and  $X_2 = 0$  represents bean media,  $X_2 = 0$  represents cucumber media.

By using R, we get following results:

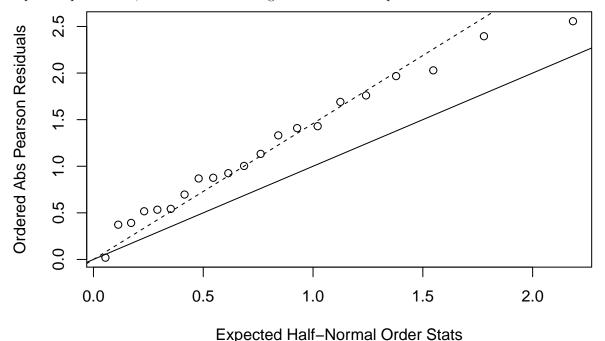
So the fitted model:

$$logit(\pi_i) = log \frac{\pi_i}{1 - \pi_i} = -0.7005 + 0.2705X_1 + 1.0647X_2$$

 $\beta_0 = -0.7005$  means the log odds of germination is 1.773 when seed is O.aegyptiaca 73 and media is bean,  $\beta_1 = 0.2705$  means the log odd ratio of germination between seed O.aegyptiaca 73 and seed O.aegyptiaca 75 is -1.738 given medias are the same and  $\beta_2 = 1.0647$  means the log odd ration of germination between bean media and cucumber media is 1.0647 given seeds are the same.

#### Question 2

When we use deviance to test goodness of fit, we got p-value=0.00230277 which is lower than 0.05 and shows the original model doesn't fit data well and is over dispersion. By calculating Pearson  $\chi^2$ , we get dispersion parameter  $\phi=2.128368$ . Following is the half normal plot:



By updating the original with the dispersion parameter, we get following results:

$$logit(\pi_i) = log \frac{\pi_i}{1 - \pi_i} = -0.7005 + 0.2705X_1 + 1.0647X_2$$

 $\beta_0 = -0.7005$  means the log odds of germination is 1.773 when seed is O.aegyptiaca 73 and media is bean,  $\beta_1 = 0.2705$  means the log odd ratio of germination between seed O.aegyptiaca 73 and seed O.aegyptiaca 75 is -1.738 given media are the same and  $\beta_2 = 1.0647$  means the log odd ration of germination between bean media and cucumber media is 1.0647 given seeds are the same.

Results show that estimated values of parameters keep the same while variances change.

## Question 3

In this problem, over dispersion might causes by dependence among each seed or difference germination rate in each seed.