

# HW3\_\_answer

Guojing Wu

2/24/2019

## Problem 1

1)

Prospective model treat the age and alcohol consumption as predictor, while disease as response. Here we propose two models and use Deviance to test which one is better:

$$Model\_less : g(\pi) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot consumption$$

$$Model\_more : g(\pi) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot consumption + \beta_3 \cdot age \cdot consumption$$

```
# 1 = (daily consumption >= 80g), 0 = (consumption < 80g)
data_eso = tibble(age = rep(c(25, 35, 45, 55, 65, 75), each = 2),
  consum = as.factor(rep(c(0, 1), 6)),
  case = c(0, 1, 5, 4, 21, 25, 34, 42, 36, 19, 8, 5),
  control = c(106, 9, 164, 26, 138, 29, 139, 27, 88, 18, 31, 0))

logit.prosp.less = glm(cbind(case, control) ~ age + consum, data = data_eso, family = binomial(link = 'logit'))
logit.prosp.more = glm(cbind(case, control) ~ age * consum, data = data_eso, family = binomial(link = 'logit'))

dif = sum(residuals(logit.prosp.less, type = "deviance")^2) - sum(residuals(logit.prosp.more, type = "deviance")^2)
pnest = 1 - pchisq(dif, 1);pnest

## [1] 0.9621536
```

The p-value we got is larger than 0.05, so we failed to reject the null hypothesis and conclude that the less model is better.

Interpretation:

- The log odds ratio of disease between exposure and nonexposure group is 0.062 for 1 unit change in age, given daily alcohol consumption fixed as a constant.
- The log odds ratio of disease between exposure and nonexposure group is 1.78 for 1 unit change in daily alcohol consumption, given age as a fixed constant.

## Problem 2

1)

This is a prospective study, treat types of seed and root extract as predictor while germination rates as response. And again, we proposed two models and use deviance analysis to test which one it better:

$$Model\_small : g(\pi) = \beta_0 + \beta_1 \cdot seed + \beta_2 \cdot root$$

$$Model\_large : g(\pi) = \beta_0 + \beta_1 \cdot seed + \beta_2 \cdot root + \beta_3 \cdot seed \cdot root$$

```

# 0 = seed 75, 1 = seed 73
# 0 = bean, 1 = cucumber
data_germ = tibble(seed = c(rep(0, 11), rep(1, 10)),
                    root = c(rep(0, 5), rep(1, 6), rep(0, 5), rep(1, 5)),
                    germ = c(c(10, 23, 23, 26, 17), c(5, 53, 55, 32, 46, 10), c(8, 10, 8, 23, 0), c(3, 2, 1, 1, 1, 1, 1, 1, 1, 1)),
                    total = c(c(39, 62, 81, 51, 39), c(6, 74, 72, 51, 79, 13), c(16, 30, 28, 45, 4), c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)))

none.disp.small = glm(cbind(germ, total - germ) ~ seed + root, data = data_germ, family = binomial(link = "logit"))
none.disp.large = glm(cbind(germ, total - germ) ~ seed * root, data = data_germ, family = binomial(link = "logit"))

dif = sum(residuals(none.disp.small, type = "deviance")^2) - sum(residuals(none.disp.large, type = "deviance")^2)
pnest = 1 - pchisq(dif, 1);pnest

```

```
## [1] 0.01136007
```

The pvalue is less than 0.05, so we reject the null hypothesis and state that the larger model is better.

Interpretation:

- The log odds ratio of germination is 0.146 between seed *O.aegyptiaca* 75 and *O.aegyptiaca* 73 while using bean root extract.
- The log odds ratio of germination is 1.318 between cucumber root extract and bean root extract while using seed *O.aegyptiaca* 75.

2)

Use goodness-of-fit test:

```
pval = 1 - pchisq(none.disp.large$deviance, 21 - 2); pval
```

```
## [1] 0.02232527
```

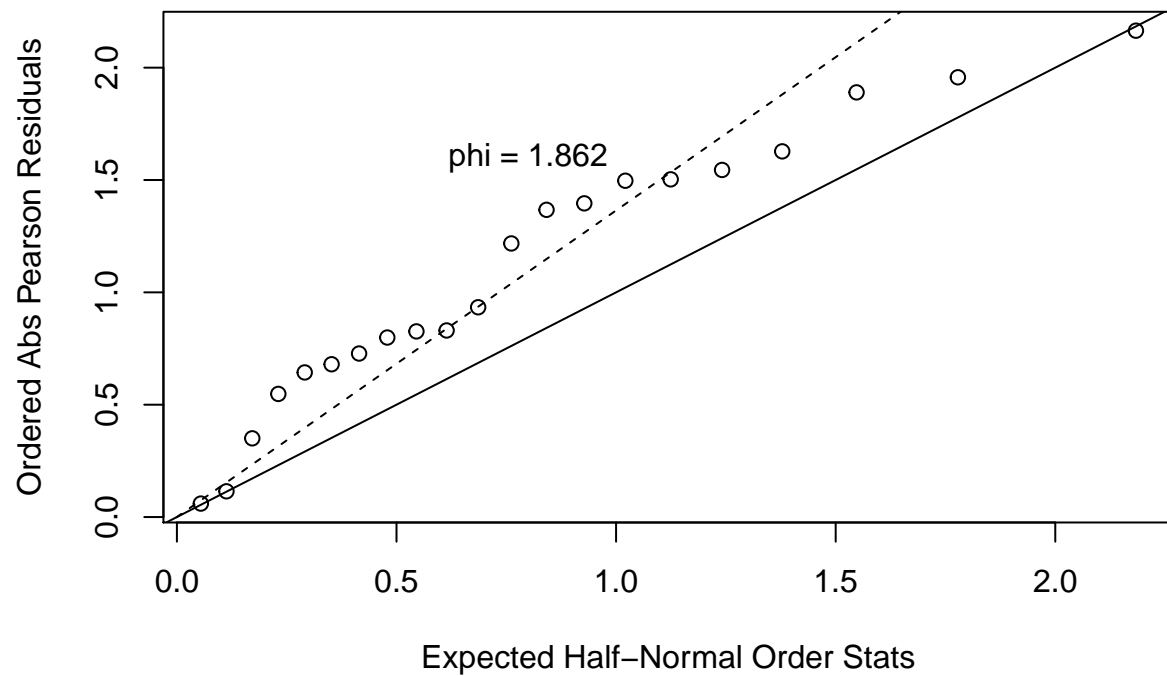
The pvalue is less than 0.05, so we reject the null hypothesis and state that this model doesn't fit the data well. Hence we confirmed that there is over dispersion. Then calculate the dispersion parameter and fit the data with new model:

```

G.stat = sum(residuals(none.disp.large, type = 'pearson') ^ 2) # pearson chisq
phi = G.stat / (21 - 4)

res = residuals(none.disp.large, type = 'pearson')
plot(qnorm((21 + 1:21 + 0.5) / (2 * 21 + 1.125)), sort(abs(res)), xlab = 'Expected Half-Normal Order Statistics')
abline(a = 0, b = 1) # it has overdispersion
abline(a = 0, b = sqrt(phi), lty = 2)
text(0.8, 1.6, paste("phi = ", round(phi, 3), sep = ""))

```



- The log odds ratio of germination is 0.146 between seed *O.aegyptiaca* 75 and *O.aegyptiaca* 73 while using bean root extract.
- The log odds ratio of germination is 1.318 between cucumber root extract and bean root extract while using seed *O.aegyptiaca* 75.

3

The possible cause of over dispersion maybe intra-class correlation