

HW3_1830

Yue Zhang

2024-10-11

3.3

(a) $P(Y=1) = 0.00109 \cdot X + 0.00255$

(b)

```
# (i) For X = 0
y1 = 0.00109 * 0 + 0.00255
y1
```

```
## [1] 0.00255
```

```
# For X = 7
y2 = 0.00109 * 7 + 0.00255
y2
```

```
## [1] 0.01018
```

```
# (ii)
RelativeRisk = y2/y1
RelativeRisk
```

```
## [1] 3.992157
```

(c)

```
X = c(0, 1, 2, 3, 4)
Y1 = c(48, 38, 5, 1, 1)
N = c(17114, 14502, 793, 127, 38)
Table1 = data.frame(X, Y1, N)
fit1 = glm(Y1/N ~ X, family = quasi(link = "identity", variance = "mu(1-mu)"),
           weights = N, data = Table1)
summary(fit1)

##
## Call:
## glm(formula = Y1/N ~ X, family = quasi(link = "identity", variance = "mu(1-mu)"),
##      data = Table1, weights = N)
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0025977 0.0005928 4.382 0.022 *
## X           0.0005044 0.0008238 0.612 0.584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasi family taken to be 2.437661)
##
## Null deviance: 6.2020  on 4  degrees of freedom
## Residual deviance: 4.9336  on 3  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 9
```

```
# Based on the re-fit model:  $P(Y=1) = 0.00050 \cdot X + 0.00260$ 
# For  $X = 0$ 
y3 = 5e-04 * 0 + 0.0026
y3
```

```
## [1] 0.0026
```

```
# For  $X = 7$ 
y4 = 5e-04 * 7 + 0.0026
y4
```

```
## [1] 0.0061
```

```
RelativeRisk2 = y4/y3
RelativeRisk2
```

```
## [1] 2.346154
```

```
# Since the relative risk changes from 3.99 to 2.35 which
# is quite dramatic, this suggests that the results are
# sensitive to the choice of scores.
```

(d)

```
# Linear model without the observation
X2 = c(0, 0.5, 1.5, 4)
Y2 = c(48, 38, 5, 1)
N2 = c(17114, 14502, 793, 127)
Table2 = data.frame(X2, Y2, N2)
fit2 = glm(Y2/N2 ~ X2, family = quasi(link = "identity", variance = "mu(1-mu)"),
           weights = N2, data = Table2)
summary(fit2)
```

```
##
```

```
## Call:
## glm(formula = Y2/N2 ~ X2, family = quasi(link = "identity", variance = "mu(1-mu)"),
##      data = Table2, weights = N2)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0026026  0.0003913   6.652  0.0219 *
## X2          0.0008184  0.0009390   0.872  0.4753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasi family taken to be 1.184679)
##
## Null deviance: 3.5131  on 3  degrees of freedom
## Residual deviance: 2.1795  on 2  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 8
```

```
# Based on the above model: P(Y=1) = 0.00082*X + 0.00260
# For X = 0
y5 = 0.00082 * 0 + 0.0026
y5
```

```
## [1] 0.0026
```

```
# For X = 7
y6 = 0.00082 * 7 + 0.0026
y6
```

```
## [1] 0.00834
```

```
RelativeRisk3 = y6/y5
RelativeRisk3
```

```
## [1] 3.207692
```

```
# Since we've calculated that the relative risk for linear
# model with the observation is 3.99, the relative risk
# decreases to 3.21 when omitting the observation.
# Therefore, the results are slightly sensitive to this
# single malformation.
```

#3.6 (a) $\log[P(Y=1)/(1-P(Y=1))]=-0.5901*X + 3.1870$ The negative coefficient for X indicates that as political ideology becomes more conservative, the likelihood of being a Democrat decreases.

(b)

```
CI_L = -0.5901 - 1.96 * 0.1564
CI_L
```

```
## [1] -0.896644
```

```
CI_U = -0.5901 + 1.96 * 0.1564
CI_U
```

```
## [1] -0.283556
```

```
# The 95% Wald CI is (-0.896644, -0.283556) while the
# profile likelihood interval is (-0.91587 -0.29832). The
# profile likelihood interval is slightly wider than the
# 95% Wald CI as it accounts for the shape of the
# likelihood function, offering a more accurate estimate.
```

(c) $Z = -3.772$ $P_value = 0.000162$ Since the p-value is less than $\alpha(0.05)$, we can reject the null hypothesis and conclude that the political ideology has a significant effect on the probability of being a Democrat.

(d)

```
Democrat = c(5, 18, 19, 25, 7, 7, 2)
N3 = c(6, 21, 20, 36, 17, 18, 3)
X3 = c(1, 2, 3, 4, 5, 6, 7)
fit3 = glm(Democrat/N3 ~ X3, family = binomial(link = "logit"),
           weights = N3)
summary(fit3)
```

```
##
## Call:
## glm(formula = Democrat/N3 ~ X3, family = binomial(link = "logit"),
##      weights = N3)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.1870      0.7002   4.552 5.33e-06 ***
## X3           -0.5901      0.1564  -3.772 0.000162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 24.7983  on 6  degrees of freedom
## Residual deviance:  7.7894  on 5  degrees of freedom
## AIC: 30.516
##
## Number of Fisher Scoring iterations: 4
```

```
fit4 = glm(Democrat/N3 ~ 1, family = binomial(link = "logit"),
           weights = N3)
summary(fit4)
```

```
##
## Call:
## glm(formula = Democrat/N3 ~ 1, family = binomial(link = "logit"),
```

```
##      weights = N3)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.7813      0.1959   3.989 6.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 24.798  on 6  degrees of freedom
## Residual deviance: 24.798  on 6  degrees of freedom
## AIC: 45.525
##
## Number of Fisher Scoring iterations: 4
```

```
lrtest(fit3, fit4)
```

```
## Likelihood ratio test
##
## Model 1: Democrat/N3 ~ X3
## Model 2: Democrat/N3 ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.258
## 2    1 -21.762 -1 17.009   3.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The test statistic(chisq) is 17.009, and p-value is
# 3.72e-05 which is much smaller than alpha(0.05).
# Therefore, we can reject the null hypothesis and state
# that X3 has a significant effect on the probability of
# being a Democrat.
```

- (e) The number of Fisher Scoring iterations is 4 indicates that the model converged efficiently, as Fisher scoring is an iterative process used to estimate the parameters of the model.

#3.8

```
Snoring = c(0, 2, 4, 5)
HeartDisease_Y = c(24, 35, 21, 30)
HeartDisease_N = c(1355, 603, 192, 224)
Table3 = data.frame(Snoring, HeartDisease_Y, HeartDisease_N)
Table3$N4 = Table3$HeartDisease_Y + Table3$HeartDisease_N
fit5 = glm(HeartDisease_Y/N4 ~ Snoring, family = binomial(link = "logit"),
           weights = N4, data = Table3)
summary(fit5)
```

```
##
## Call:
## glm(formula = HeartDisease_Y/N4 ~ Snoring, family = binomial(link = "logit"),
##      data = Table3, weights = N4)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.86625    0.16621 -23.261  < 2e-16 ***
## Snoring      0.39734    0.05001   7.945 1.94e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 65.9045  on 3  degrees of freedom
## Residual deviance:  2.8089  on 2  degrees of freedom
## AIC: 27.061
##
## Number of Fisher Scoring iterations: 4
```

```
confint(fit5)
```

```
##             2.5 %      97.5 %
## (Intercept) -4.2072190 -3.5544117
## Snoring      0.2999362  0.4963887
```

#3.13 (a)

```
Crab = read.table("http://www.stat.ufl.edu/~aa/cat/data/Crabs.dat",
  header = TRUE)
fit6 = glm(sat ~ weight, family = poisson(link = "log"), data = Crab)
summary(fit6)
```

```
##
## Call:
## glm(formula = sat ~ weight, family = poisson(link = "log"), data = Crab)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.42841    0.17893  -2.394   0.0167 *
## weight       0.58930    0.06502   9.064  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##    Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 560.87  on 171  degrees of freedom
## AIC: 920.16
##
## Number of Fisher Scoring iterations: 5
```

```
# Based on above, the prediction equation is: mu_hat =
# exp(0.58930*X - 0.42841) For x = 2.44kg,
MeanResponse = exp(0.5893 * 2.44 - 0.42841)
MeanResponse
```

```
## [1] 2.744179
```

(b)

```
exp(0.5893)
```

```
## [1] 1.802726
```

```
# The beta hat means that for every 1kg increase in crab  
# weight, the expected number of satellites increases by  
# approximately 80.3%.
```

```
CI_L2 = 0.5893 - 1.96 * 0.06502  
CI_L2
```

```
## [1] 0.4618608
```

```
CI_U2 = 0.5893 + 1.96 * 0.06502  
CI_U2
```

```
## [1] 0.7167392
```

```
# The 95% CI is (0.461861, 0.716739).
```

```
exp(0.461861)
```

```
## [1] 1.587025
```

```
exp(0.716739)
```

```
## [1] 2.047745
```

```
# The 95% CI for multiplicative effect is (1.58703,  
# 2.04775)
```

(c)

```
# Wald test Null hypothesis: beta = 0, Alternative  
# hypothesis: beta is not equal to 0.  
Z_squared = (0.5893/0.06502)^2  
Z_squared
```

```
## [1] 82.14459
```

```
# Likelihood Ratio test  
fit7 = glm(sat ~ 1, family = poisson(link = "log"), data = Crab)  
summary(fit7)
```

```
##
## Call:
## glm(formula = sat ~ 1, family = poisson(link = "log"), data = Crab)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.0713      0.0445   24.07  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 632.79  on 172  degrees of freedom
## AIC: 990.09
##
## Number of Fisher Scoring iterations: 5
```

```
lrtest(fit6, fit7)
```

```
## Likelihood ratio test
##
## Model 1: sat ~ weight
## Model 2: sat ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    2 -458.08
## 2    1 -494.04 -1 71.925  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Since the Likelihood ratio test statistic is 71.925 and
# df = 1, and the p-value is far less than 0.05 which means
# that we can reject the null hypothesis.
```