Hw8_Q4

Hongkai Lou

2024-12-03

```
dat = expand.grid(factory=c("East", "West"), accident=c("No", "Yes"))
dat$y = c(645,1275, 28,31)
tab = matrix(dat$y, nrow=2,
dimnames=list(factory=c("East", "West"), accident=c("No", "Yes")))
```

4(a)

```
fit1 = glm(y~factory+accident, poisson, dat)
summary(fit1)
```

```
##
## Call:
## glm(formula = y ~ factory + accident, family = poisson, data = dat)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.48148    0.03875  167.27   <2e-16 ***
## factoryWest  0.66298    0.04745   13.97   <2e-16 ***
## accidentYes -3.48254    0.13217  -26.35   <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: 2423.492  on 3  degrees of freedom
## Residual deviance:    4.678  on 1  degrees of freedom
## AIC: 38.427
##
## Number of Fisher Scoring iterations: 4
```

```
print(paste('The number of unlogged accidents in the west is', exp(6.481+0.663-3.483)))
```

```
## [1] "The number of unlogged accidents in the west is 38.9002236513537"
```

```
fit2 = glm(y~factory*accident, poisson, dat)
summary(fit2)
```

```
##
## Call:
## glm(formula = y ~ factory * accident, family = poisson, data = dat)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)              6.46925    0.03937 164.299   <2e-16 ***
## factoryWest              0.68145    0.04832  14.103   <2e-16 ***
## accidentYes             -3.13705    0.19304 -16.251   <2e-16 ***
## factoryWest:accidentYes -0.57967    0.26515  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2.4235e+03  on 3  degrees of freedom
## Residual deviance: 1.1857e-13  on 0  degrees of freedom
## AIC: 35.749
##
## Number of Fisher Scoring iterations: 3
```

```
print(paste('The number of unlogged accidents in the west using interaction model is', exp(6.46925+0.68
```

```
## [1] "The number of unlogged accidents in the west using interaction model is 30.999776661765"
```

d

The residual deviance 0 residual deviance largely due to 0 degrees of freedom. The degrees of freedom of chi-square test is (r-1)* (c-1). Here, the degrees of freedom is 1*1, but with intersection term included, it becomes 0, meaning the model fully explains the data, and the chi-square tests lost its significance.

e

We see the z value is -2.186 with p-value equals to 0.0288. Since p-value < 0.05, it is considered significant.

f

By rejecting the null hypothesis, it tells us that we reject the null hypothesis that the independent variable will significantly affect the output variable. Thus, Factory and Accident are not independent.

g

```
lrt_result <- anova(fit1, fit2, test = "Chisq")

print(lrt_result)
```

```
## Analysis of Deviance Table
##
## Model 1: y ~ factory + accident
## Model 2: y ~ factory * accident
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         1      4.678
## 2         0      0.000  1    4.678  0.03055 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The deviance is 4.678 with a p value of 0.03055, which is larger than 0.05, so we reject the null hypothesis

h

```
#Factory: i, accident: j
summary(fit2)
```

```
##
## Call:
## glm(formula = y ~ factory * accident, family = poisson, data = dat)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               6.46925    0.03937 164.299   <2e-16 ***
## factoryWest               0.68145    0.04832  14.103   <2e-16 ***
## accidentYes              -3.13705    0.19304 -16.251   <2e-16 ***
## factoryWest:accidentYes -0.57967    0.26515  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2.4235e+03  on 3  degrees of freedom
## Residual deviance: 1.1857e-13  on 0  degrees of freedom
## AIC: 35.749
##
## Number of Fisher Scoring iterations: 3
```

```
east_noacc = exp(6.46925)
east_acc = exp(6.46925-3.13705)
west_noacc = exp(6.46925+0.68145)
west_acc = exp(6.46925+0.68145-3.13705-0.57967)
east_accident_odds = log(east_acc/east_noacc)
west_accident_odds = log(west_acc/west_noacc)
print(c(east_accident_odds, west_accident_odds))
```

```
## [1] -3.13705 -3.71672
```

i

c = -3.13705, d = (3.13705-3.71672) = -0.57967

j

The estimation is exactly same

```
summary(glm(accident~factory, family = 'binomial', weight = y, dat))
```

```
##
## Call:
## glm(formula = accident ~ factory, family = "binomial", data = dat,
##     weights = y)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.1370     0.1930 -16.251   <2e-16 ***
```

```
## factoryWest  -0.5797     0.2651  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 530.73  on 3  degrees of freedom
## Residual deviance: 526.06  on 2  degrees of freedom
## AIC: 530.06
##
## Number of Fisher Scoring iterations: 6
```