Applied Statistics - MATH 7343

Sai Nikhil
NUID: 002564864

20.5.8

(1)

(a)

```
> data <- read.table(file="stenosis.txt", header = TRUE)
>
> lr.fit <- glm(data$disease~data$smoke, data = data, family = binomial(link = "logit"))
> summary(lr.fit)

Call:
glm(formula = data$disease ~ data$smoke, family = binomial(link = "logit"),
    data = data)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.251  -1.087  -1.087   1.188   1.270

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2157     0.1829  -1.180    0.238
data$smoke    0.3863     0.2762   1.399    0.162

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 297.94  on 214  degrees of freedom
Residual deviance: 295.97  on 213  degrees of freedom
AIC: 299.97

Number of Fisher Scoring iterations: 3

>
```

$$\therefore \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.2157 + 0.3863 x_1$$

$0.3863 > 0 \Rightarrow$ log odds of developing aortic stenonis

$\therefore$ p is higher for individuals who smoke relative to those who don't. Further, it implies that with smoking the log odds of developing aortic stenosis increases by 0.3863. Also, log odds increases as "$\hat{p}$" increases.

(b)

```
> exp(0.3863)
[1] 1.471526
```

estimated odds of suffering from aortic stenosis for individuals who smoke relative to those who do not is $\widehat{OR} = 1.47$

(c)

```
> exp(0.3863 + c(-1, 1) * 1.96 * 0.2762)
[1] 0.8563706 2.5285651
```

95% CI $\rightarrow (0.856, 2.529)$

It contains "1" $\Rightarrow$ there is no evidence that probability of developing aortic stenosis is different depending on smoking status. Since "1" is in the interval, we can conclude that 1.47 is not different from "1".

(d)

```
> lr.fit2 <- glm(data$disease ~ data$smoke + data$sex, data = data, family = binomial(link = "logit"))
> summary(lr.fit2)

Call:
glm(formula = data$disease ~ data$smoke + data$sex, family = binomial(link = "logit"),
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3630  -1.0555  -0.9783   1.0807   1.3905

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4882     0.2159  -2.261   0.0238 *
data$smoke    0.1946     0.2903   0.670   0.5026
data$sex      0.7199     0.2881   2.499   0.0125 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 297.94  on 214  degrees of freedom
Residual deviance: 289.64  on 212  degrees of freedom
AIC: 295.64

Number of Fisher Scoring iterations: 4


> exp(0.1946)
[1] 1.214825
>
```

$\widehat{OR} = 1.215$. This is little lower then unadjusted for gender.

(e)

```
> exp(0.1946 + c(-1, 1) * 1.96 * 0.2903)
[1] 0.687710 2.145962
```

95% CI → (0.688, 2.146)

It Contains "1". Since "1" is in the interval we can conclude that 1.21 is not different from "1".

(f) In order to determine if the relationship between the presence of aortic stenotic and smoking status differs for males and females, we need to include in the logistic regression model an interaction term that is the product of the two dichotomous explanatory variables.