Homework 9

Yuanyou Yao

2020/4/23

Problem 1

(a)

The parameter of defendant's is negative while the parameter of victim's is positive. Therefore we can conclude that black defendants for killing white victims are more likely to have the "yes".\
The probability is

$$\hat{\pi} = \exp(-3.60 + 2.40) = 0.23.$$

(b)

The parameter for victim's race means that the conditional log-odds ratio of receiving the death penalty for white against black victims is 2.40. The conditional odds ratio is $\exp(2.40) = 11.07$, meaning that conditioning on defendant race, the odds of receiving the death penalty are 11 times highter in white victims group.

(c)

$$(\exp(1.3068), \exp(3.7175)) = (3.69, 41.16).$$

(d)

As the output shown in the table, the likelihood ratio test statistic is 20.35, associating with the p-value $p < .0001$. So we reject the null hypothesis that victims' race is independent to death penalty.

Problem 2

(a)

```r
Success=c(11,10,16,22,14,7,2,1,6,0,1,0,1,1,4,6)
Failure=c(25,27,4,10,5,12,14,16,11,12,10,10,4,8,2,1)
Treatment=rep(c(1,0),8)
Center2=c(0,0,1,1,rep(0,12))
Center3=c(rep(0,4),1,1,rep(0,10))
Center4=c(rep(0,6),1,1,rep(0,8))
Center5=c(rep(0,8),1,1,rep(0,6))
Center6=c(rep(0,10),1,1,rep(0,4))
Center7=c(rep(0,12),1,1,rep(0,2))
Center8=c(rep(0,14),1,1)
Center=cbind(Center2,Center3,Center4,Center5,Center6,Center7,Center8)
summary(glm(cbind(Success,Failure)~Treatment+Center,family = binomial(link = 'logit')))
```

```
##
## Call:
## glm(formula = cbind(Success, Failure) ~ Treatment + Center, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.87919  -0.77729  -0.00401   0.47293   0.92934
```

```
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.3220     0.3165  -4.177 2.95e-05 ***
## Treatment       0.7769     0.3067   2.533  0.01130 *
## CenterCenter2   2.0554     0.4201   4.893 9.94e-07 ***
## CenterCenter3   1.1529     0.4246   2.715  0.00662 **
## CenterCenter4  -1.4185     0.6636  -2.138  0.03255 *
## CenterCenter5  -0.5199     0.5338  -0.974  0.33007
## CenterCenter6  -2.1469     1.0614  -2.023  0.04310 *
## CenterCenter7  -0.7977     0.8149  -0.979  0.32764
## CenterCenter8   2.2079     0.7195   3.069  0.00215 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 93.5545  on 15  degrees of freedom
## Residual deviance:  9.7463  on  7  degrees of freedom
## AIC: 66.136
##
## Number of Fisher Scoring iterations: 4
```

We notice that the probablity of success vary from center to center. The odds ratios of success between drug and control are held fixed within each center.

(b)

The reason is that we want to fit the data without the treatment variable and compare. The chi-square statistic is $90.96 - 9.7463 = 81.2137$ on 7 d.f. This is significant and indicating that treatment center having an association with the probability of success.

Problem 3

(a)

Conditional on religious affiliation and political party affiliation, the odds of female are $\exp(0.16) = 1.17$ times higher than that of male.

(b)

$$log(\hat{\pi}_1) = -0.11 + 0 - 0.66 - 1.67 = -2.44,$$
$$\hat{\pi}_1 = \frac{\exp(-2.44)}{1+\exp(-2.44)} = 0.08,$$
$$log(\hat{\pi}_2) = -0.11 + 0.16 + 0 + 0.84 = 0.89,$$
$$\hat{\pi}_2 = \frac{\exp(0.89)}{1+\exp(0.89)} = 0.71.$$

(c)

Since the differences between categorical variable doesn't change, we obtain $\hat{\beta}_1^G = 0$ and $\hat{\beta}_2^G = -0.16$.\
According to (a), the odds of a male supports laws legalizing abortion are $1/1.17 = 0.85$ times the odds of female.

(d)

Since the differences between categorical variable doesn't change, we obtain $\hat{\beta}_1^G = 0.08$ and $\hat{\beta}_2^G = -0.08$.\
The odds ratio are still the same.

Problem 4

(a)

When $R = 1$ and $R = 0$, the prediction equations are

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -6.70 + 0.10 \times \text{Alcohol} + 1.40 \times \text{Smoking},$$

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -7.00 + 0.10 \times \text{Alcohol} + 1.20 \times \text{Smoking},$$

respectively.\
Thus, the conditional YS odds ratio are $\exp(1.40) = 4.05$ and $\exp(1.20) = 3.32$ respectively.\
Similarly, when $S = 1$ and $S = 0$, the prediction equations are

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -5.80 + 0.10 \times \text{Alcohol} + 0.50 \times \text{Race},$$

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -7.00 + 0.10 \times \text{Alcohol} + 0.30 \times \text{Race},$$

respectively.\
Thus, the conditional YR odds ratio are $\exp(0.50) = 1.65$ and $\exp(0.30) = 1.35$ respectively.\

(b)

The coefficient of R represents the log odds ratio of cancer for black non-smokers against white non-smoker. The coefficient of S represents the log odds ratio of cancer for white smokers against white non-smokers.\
The hypothesis for R is that race has no association with cancer. The hypothesis for S is that smoking has no association with cancer.

(c)

The difference between the effect of A for blacks and for whites is

$$0.04 - 0 = 0.04,$$

this means that the effect of will increase 0.04 comparing with whites.

Problem 5

(a)

```r
d=c(45,15,40,83,90,25,35,65,95,35,75,45,50,75,30,25,20,60,70,30,60,61,65,15,20,45,15,25,15,30,40,15,135
t=c(0,0,0,1,1,1,0,0,0,0,0,1,1,1,0,0,1,1,1,0,0,0,0,1,1,0,1,0,1,0,0,1,1,1,1)
y=c(0,0,1,1,1,1,1,1,1,1,1,1,0,1,0,1,0,1,1,1,1,0,1,0,0,1,0,1,0,1,1,0,1,0,0)
summary(glm(y~t+d,family = binomial(link = 'logit')))
```

```
##
## Call:
## glm(formula = y ~ t + d, family = binomial(link = "logit"))
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -2.3802  -0.5358   0.3047   0.7308   1.7821
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.41734    1.09457  -1.295  0.19536
## t           -1.65895    0.92285  -1.798  0.07224 .
## d            0.06868    0.02641   2.600  0.00931 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 46.180  on 34  degrees of freedom
## Residual deviance: 30.138  on 32  degrees of freedom
## AIC: 36.138
##
## Number of Fisher Scoring iterations: 5
```

The odds of sore throat increase $\exp(0.06868) = 1.07$ times for surgery duration increasing. As for T, the odds of sore threat decrease to $\exp(-1.67895) = 0.19$.

(b)

```
library(lmtest)
lrtest(lm(y~t+d,family = binomial(link = 'logit')),lm(y~t,family = binomial(link = 'logit')))
```

```
## Likelihood ratio test
##
## Model 1: y ~ t + d
## Model 2: y ~ t
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   4 -16.195
## 2   3 -22.343 -1 12.296  0.0004539 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the result shows, the p-value is 0.0004539. So we have to reject the reduced model.

(c)

The prediction equations are

$$\text{logit}(\hat{\pi}) = -4.42 + 0.10d,$$

$$\text{logit}(\hat{\pi}) = 0.05 + 0.03d,$$

respectively.\
For (i), odds of sore throat are $\exp(0.10) = 1.11$ times when surgery time increases. For (ii), laryngeal mask increases by 3%. Therefore, odds of sore throat increase 8% more in tracheal tube comparing to laryngeal mask.

(d)

```
lrtest(lm(y~t+d,family = binomial(link = 'logit')),lm(y~t*d,family = binomial(link = 'logit')))
```

```
## Likelihood ratio test
##
## Model 1: y ~ t + d
```

```
## Model 2: y ~ t * d
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   4 -16.195
## 2   5 -15.353  1 1.6847     0.1943
```

The result tells us that we cannot reject the null hypothesis. So the model without interaction fits well.