

## Assignment 2

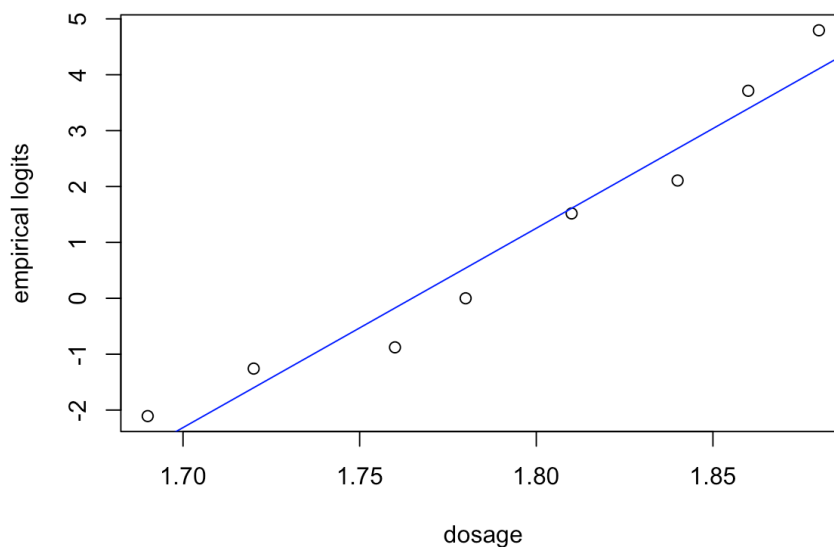
QINGTAN SHEN  
1130945

1.(a)

First we preprocessed the data, then we fit a linear model for empirical logit and dosage, here is the R code:

```
#1
#1.(a)
dosage = c(1.69,1.72,1.76,1.78,1.81,1.84,1.86,1.88)
nb = c(59,60,62,56,63,59,62,60)
nk = c(6,13,18,28,52,53,61,60)
em_logit = log((nk+0.5) / (nb-nk+0.5))
reg = lm(em_logit ~ dosage)
plot(dosage, em_logit, xlab = 'dosage', ylab = 'empirical logits')
abline(reg, col='blue')
```

And here is the plot:



The plot looks linear, but not significant.

1.(b)

```
> #1.(b)
> model1 <- glm(nk/nb ~ dosage, family = binomial, weights = nb)
> model1$coefficients
(Intercept)      dosage
  -60.10328    33.93416
```

We fit this logistic regression model, and the intercept is -60.103, the dosage is 33.934

1.(c)

$$interval = \mu \pm z_{\alpha/2} * \sigma$$

$$\mu = 33.934, \quad \sigma = 2.903, \quad z_{\alpha/2} = 1.96$$

So here is the code:

```
> b_interval = c(33.934-1.96*2.903, 33.934+1.96*2.903)
> b_interval
[1] 28.24412 39.62388
```

Our interval is [28.244, 39.624].

1.(d)

$$\beta_0 + \beta_1 * x = \ln\left(\frac{0.5}{1-0.5}\right) = 0$$

$$x = -\frac{\beta_0}{\beta_1} = -\frac{60.103}{33.934} = 1.771$$

The value for dosage is 1.771

1.(e)

Here is the formula:

$$\ln(o1) = \ln\left(\frac{p1}{1-p1}\right) = \beta_0 + \beta_1 * x$$

$$\ln(o2) = \ln\left(\frac{p2}{1-p2}\right) = \beta_0 + \beta_1 * (x + 0.1)$$

$$\ln(o2) - \ln(o1) = \ln\left(\frac{o2}{o1}\right) = \beta_1 * 0.1$$

$$\frac{o2}{o1} = e^{0.1 * \beta_1}$$

$$0.1 * \beta_1 \in 0.1 * [28.244, 39.624] = [2.824, 3.962]$$

$$\frac{o2}{o1} \in [e^{2.824}, e^{3.962}] = [16.844, 52.562]$$

The confidence interval is [16.844, 52.562]

1.(f)

```
> summary(model1)
```

Call:

```
glm(formula = nk/nb ~ dosage, family = binomial, weights = nb)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.8986	-0.5475	0.9842	1.3315	1.7179

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-60.103	5.164	-11.64	<2e-16 ***
dosage	33.934	2.903	11.69	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 284.202 on 7 degrees of freedom  
Residual deviance: 13.633 on 6 degrees of freedom  
AIC: 43.831

Number of Fisher Scoring iterations: 4

We get variance information from this summary operation.

$$\text{var}(\beta_0 + \beta_1 * x) = \text{var}(\beta_0) + x^2 \text{var}(\beta_1) + 2x * \rho * \sqrt{\text{var}(\beta_0) * \text{var}(\beta_1)}$$

$$x = 1.8, \rho = -1, \sqrt{\text{var}(\beta_0)} = 5.164164, \sqrt{\text{var}(\beta_1)} = 2.902867$$

$$\text{So: } \text{var}(\beta_0 + \beta_1 * 1.8) = 0.0208$$

$$\mu(\beta_0 + \beta_1 * 1.8) = -60.103 + 1.8 * 33.934 = 0.9782$$

$$p_{\mu} = \frac{e^{0.9782}}{1 + e^{0.9782}} = 0.727$$

$$\begin{aligned} \log_{\text{interval}} &= [0.9782 - 1.96 * \sqrt{0.0208}, 0.9782 + 1.96 * \sqrt{0.0208}] \\ &= [0.696, 1.261] \end{aligned}$$

$$p_{\text{interval}} = \left[ \frac{e^{0.696}}{1 + e^{0.696}}, \frac{e^{1.261}}{1 + e^{1.261}} \right] = [0.667, 0.779]$$

The estimate value is 0.727, the confidence interval is [0.667, 0.779].

1.(g)

First part: we use residual deviance

```
> #1.(g)
> # residual deviance
> d1 = deviance(model1)
> df1 = 8 - 2
> 1 - pchisq(d1, df=df1)
[1] 0.03401062
```

We got the p-value for model1 is 0.034, which is smaller than 0.05, so the model is not adequate at 95%.

Second part: we use Pearson test

```
> # pearson test  
> pr = sum(resid(model1,type="pearson")^2)  
> 1 - pchisq(pr, df=df1)  
[1] 0.05948877
```

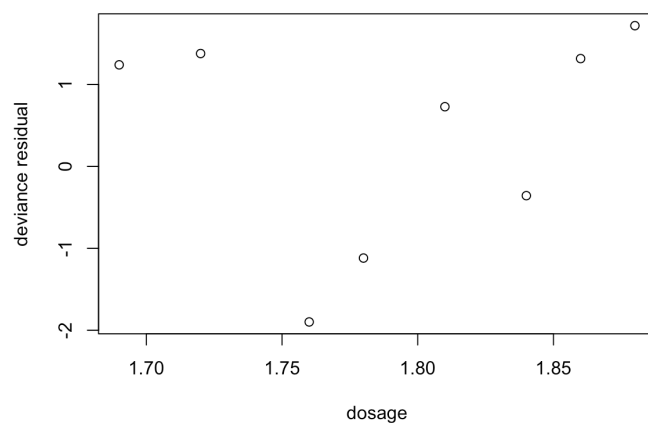
The p-value for Pearson test is 0.0595, which is a little larger than 0.05. Although we can say that this model is adequate at 95% level, we are not quite sure.

1.(h)

Here is the plot code:

```
> #1.(h)  
> res <- resid(model1)  
> plot(dosage, res, xlab='dosage', ylab='deviance residual')
```

Here is the plot:



From this plot we can see that the deviance residual may not be random. Especially when dosage is greater than 1.80, the residual is increased more obviously. Because the number of points is too small, we cannot find a clear trend from the whole graph.

1.(i)

Here is the model we create:

```

> #1.(i)
> model2 <- glm(nk/nb ~ dosage + I(dosage^2), family = binomial, weights
= nb)
> summary(model2)

Call:
glm(formula = nk/nb ~ dosage + I(dosage^2), family = binomial,
weights = nb)

Deviance Residuals:
    1      2      3      4      5      6
-0.44733  0.97664 -0.88177 -0.03034  1.01747 -1.17598
    7      8
 0.24516  0.83445

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   422.42    171.99   2.456  0.01405 *
dosage        -510.65    194.80  -2.621  0.00876 **
I(dosage^2)    153.56     55.14   2.785  0.00535 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:   5.107  on 5  degrees of freedom
AIC: 37.305

Number of Fisher Scoring iterations: 4

```

First, we compare the deviance residual with model1, here is the code:

```

> 1 - pchisq(5.107, df=8-3)
[1] 0.4029619

```

p-value is 0.403, much better than 0.034 in model1, so model2 is more adequate.

Second, we use Pearson test on model2, here is the code:

```

> pr1 = sum(resid(model2,type="pearson")^2)
> 1 - pchisq(pr1, df=8-3)
[1] 0.4175816

```

p-value for Pearson test is 0.418, which is greater than 0.059 in model1, so model2 is better.

Finally, we use anova test to find whether  $I(\text{dosage}^2)$  is significant, here is the code:

```

> anova(model1, model2, test='Chi')
Analysis of Deviance Table

Model 1: nk/nb ~ dosage
Model 2: nk/nb ~ dosage + I(dosage^2)
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         6    13.633
2         5     5.107  1   8.5264   0.0035 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

P-value for  $I(\text{dosage}^2)$  is 0.0035, so this variable is significant.

In conclusion, quadratic logistic regression model provides a better fit to the data than the straight-line model.

2.(a)

First, we preprocess the data, here is the code:

```
> #2.(a)
> education <- factor(rep(6:17, 2))
> # male as 0, female as 1
> gender <- factor(c(rep(0, 12), rep(1, 12)))
> agree <- c(25,27,75,29,32,36,115,31,28,9,15,3,17,26,91,30,55,50,190,17,
18,7,13,3)
> disagree <- c(9,15,49,29,45,59,245,70,79,23,110,29,5,16,36,35,67,62,40
3,92,81,34,115,28)
> total <- agree + disagree
```

Then we create model.a1 and use anova to test it:

```
> model.a1 <- glm(agree/total ~ education*gender, family=binomial, weights=total)
> anova(model.a1, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit
Response: agree/total

Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			23	337.90	
education	11	322.62	12	15.28	<2e-16 ***
gender	1	0.12	11	15.16	0.7298
education:gender	11	15.16	0	0.00	0.1753

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

After that, we create the same model with different order model.a2

```
> model.a2 <- glm(agree/total ~ gender*education, family=binomial, weights=total)
> anova(model.a2, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit
Response: agree/total

Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			23	337.90	
gender	1	0.37	22	337.53	0.5431
education	11	322.37	11	15.16	<2e-16 ***
gender:education	11	15.16	0	0.00	0.1753

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

As we see from the model.a1 and model.a2, the p-value for interaction of gender and education is 0.175, which is larger than 0.05, so we can ignore it.

After that, we create model.a3, without considering the interaction between gender and education:

```
> model.a3 <- glm(agree/total ~ education + gender, family=binomial, weights=total)
> anova(model.a3, test='Chi')
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: agree/total

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			23	337.90	
education	11	322.62	12	15.28	<2e-16 ***
gender	1	0.12	11	15.16	0.7298

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

After that, we create model.a4, just change the variable order in model.a3

```
> model.a4 <- glm(agree/total ~ gender + education, family=binomial, weights=total)
> anova(model.a4, test='Chi')
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: agree/total

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			23	337.90	
gender	1	0.37	22	337.53	0.5431
education	11	322.37	11	15.16	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

p-value for gender variable is 0.5431, which is greater than 0.05, so gender is not significant.

So we create model.a5 as our final model, ignoring the gender variable

```
> model.a5 <- glm(agree/total ~ education, family=binomial, weights=total)
> summary(model.a5)
```

Call:

```
glm(formula = agree/total ~ education, family = binomial, weights = total)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.88558	-0.34518	-0.00269	0.31440	1.81107

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.0986	0.3086	3.560	0.000371 ***
education7	-0.5623	0.3826	-1.470	0.141617
education8	-0.4293	0.3362	-1.277	0.201649
education9	-1.1800	0.3575	-3.301	0.000965 ***
education10	-1.3512	0.3401	-3.973	7.09e-05 ***
education11	-1.4401	0.3393	-4.244	2.19e-05 ***
education12	-1.8522	0.3163	-5.855	4.76e-09 ***
education13	-2.3150	0.3496	-6.621	3.56e-11 ***
education14	-2.3451	0.3510	-6.681	2.38e-11 ***
education15	-2.3691	0.4187	-5.659	1.53e-08 ***
education16	-3.1825	0.3680	-8.649	< 2e-16 ***
education17	-3.3499	0.5286	-6.337	2.34e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 337.903 on 23 degrees of freedom  
 Residual deviance: 15.279 on 12 degrees of freedom  
 AIC: 147.73

## 2.(b)

First, we change education variable from factor to number, then we create model.b:

```
> education.num <- c(rep(6:17, 2))
> model.b <- glm(agree/total ~ (education.num + gender)^2, family=binomial, weights=total)
> summary(model.b)
```

Call:

```
glm(formula = agree/total ~ (education.num + gender)^2, family = binomial,
     weights = total)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.20884	-0.45794	-0.07201	0.51043	1.77254

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.45543	0.29343	8.368	<2e-16 ***
education.num	-0.26352	0.02527	-10.426	<2e-16 ***
gender1	0.89268	0.43041	2.074	0.0381 *
education.num:gender1	-0.08172	0.03720	-2.197	0.0280 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 337.903 on 23 degrees of freedom  
 Residual deviance: 20.244 on 20 degrees of freedom  
 AIC: 136.69

Then we use anova test to model.b:



```
> anova(model.b, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: agree/total

Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			23	337.90	
education.num	1	312.666	22	25.24	< 2e-16 ***
gender	1	0.149	21	25.09	0.69948
education.num:gender	1	4.843	20	20.24	0.02776 *

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that although p-value for gender variable is 0.699, greater than 0.05, but we cannot ignore it because the p-value of interaction between education and gender is 0.028, which is significant. So we cannot ignore any variables. Our final model is model.b.

Here comes to the interpretation part:

First, if education added 1 for male:

```
> exp(1 * (-0.26352))
[1] 0.7683423
```

For male, if education added 1, the odds ratio will be 76.83% of the original odds ratio.

Second, if education added 1 for female:

```
> exp(1 * (-0.26352) + 1 * (-0.08172))
[1] 0.7080504
```

For female, if education added 1, the odds ratio will be 70.81% of the original odds ratio.

Third, we consider the odds ratio for females against males in the same education level:

$$\text{logit} = 0.89268 - 0.08172 * \text{education\_level}$$

```
> c1 <- c(rep(6:17))
> c2 <- exp(-0.08172 * c1 + 0.89268)
> c2
[1] 1.4953496 1.3780094 1.2698769 1.1702296 1.0784017 0.9937794 0.9157975 0.8439348
[9] 0.7777112 0.7166842 0.6604459 0.6086207
```

Here is the table:

Education level	Odds ratio for females against males
6	1.495
7	1.378
8	1.270
9	1.170
10	1.078
11	0.994
12	0.916
13	0.844
14	0.778
15	0.717
16	0.660
17	0.609

2.(c)

First we create a new model.t which has both factor education and numerical education

```
> model.t <- glm(agree/total ~ education.num + education.num:gender + gender + education,
  family=binomial, weights=total)
> anova(model.t, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit
Response: agree/total

Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                  23    337.90
education.num         1   312.666    22    25.24 < 2e-16 ***
gender                1    0.149    21    25.09 0.69948
education            10    9.927    11    15.16 0.44690
education.num:gender  1    5.213    10    9.95 0.02242 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that the p-value for factor of education is 0.447, which is not significant.

Then we compare model.t with model.b:

```
> anova(model.b, model.t, test='Chi')
Analysis of Deviance Table

Model 1: agree/total ~ (education.num + gender)^2
Model 2: agree/total ~ education.num + education.num:gender + gender +
  education
      Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          20    20.2442
2           10    9.9473 10    10.297 0.4148
```

The p-value for these ten factor education variables is 0.4148, also shows that they are not significant, and can be ignored.

We can also analyze the deviance residual of model.b:

```
> 1 - pchisq(deviance(model.b), df=20)
[1] 0.4427514
```

The p-value of deviance residual for model.b is 0.443, which is greater than 0.05, showing that model.b is adequate.

Also we can get more information about the interaction between gender and education level in model.b, it shows this interaction is significant to people's attitude, however we can only get the relationship between people's attitude and their education level in model.a2, which may not be enough. Also model.b is simple to understand.

In conclusion, model.b is better.

3.(a)

First we preprocess the data:

```
> #3.(a)
> # white is 1, black is 0
> DR <- c(1,1,1,1,0,0,0,0)
> VR <- c(1,1,0,0,1,1,0,0)
> # yes is 1, no is 0
> DP <- c(1,0,1,0,1,0,1,0)
> value <- c(19,132,0,9,11,52,6,97)
```

For (a).i, we create model.c1 for this part, here is the code:

```
> # (a).i
> model.c1 <- glm(value ~ DR + VR + DP, family=poisson)
> summary(model.c1)

Call:
glm(formula = value ~ DR + VR + DP, family = poisson)

Deviance Residuals:
    1     2     3     4     5     6     7     8 
1.9881  3.7542 -3.4843 -7.0237 -0.3023 -5.0100 -0.1196  5.7623 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.92657    0.11075  35.455 < 2e-16 ***
DR           -0.03681    0.11079  -0.332    0.74
VR            0.64748    0.11662   5.552 2.83e-08 ***
DP           -2.08636    0.17671 -11.807 < 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: 395.92  on 7  degrees of freedom
Residual deviance: 137.93  on 4  degrees of freedom
AIC: 181.61
```

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(137.93, df=4)
[1] 0
```

The value we calculated is almost 0, which is smaller than 0.05, so it is not a adequate model at 95%, so we reject that all three factors are mutually independent.

For (a).ii, we create model.c2 for this part, here is the code:

```
> #(a).ii
> model.c2 <- glm(value ~ DP + DR * VR, family=poisson)
> summary(model.c2)
```

Call:  
glm(formula = value ~ DP + DR \* VR, family = poisson)

Deviance Residuals:

1	2	3	4	5	6	7	8
0.5569	-0.2012	-1.4099	0.3443	1.4118	-0.5467	-1.7531	0.5561

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.5177	0.1004	44.976	< 2e-16 ***
DP	-2.0864	0.1767	-11.807	< 2e-16 ***
DR	-2.4375	0.3476	-7.013	2.34e-12 ***
VR	-0.4916	0.1599	-3.074	0.00212 **
DR:VR	3.3116	0.3786	8.748	< 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: 395.9153 on 7 degrees of freedom  
Residual deviance: 8.1316 on 3 degrees of freedom  
AIC: 53.813

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(8.1316, df=3)
[1] 0.0433688
```

The value we calculated is 0.043, which is smaller than 0.05, so it is not a adequate model at 95%, so we reject that sentence is independent of both the defendant's and the victim's race.

For (a).iii, we create model.c3 for this part, here is the code:

```
> #(a).iii
> model.c3 <- glm(value ~ DR * DP + DR * VR, family=poisson)
> summary(model.c3)
```

Call:  
glm(formula = value ~ DR \* DP + DR \* VR, family = poisson)

Deviance Residuals:

1	2	3	4	5	6	7	8
0.24994	-0.09277	-1.46202	0.37142	1.62523	-0.61322	-1.52514	0.46922

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.5267	0.1020	44.396	< 2e-16 ***
DR	-2.4559	0.3498	-7.021	2.2e-12 ***
DP	-2.1707	0.2560	-8.480	< 2e-16 ***
VR	-0.4916	0.1599	-3.074	0.00212 **
DR:DP	0.1664	0.3539	0.470	0.63821
DR:VR	3.3116	0.3786	8.748	< 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: 395.9153 on 7 degrees of freedom  
Residual deviance: 7.9102 on 2 degrees of freedom  
AIC: 55.592

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(7.9102, df=2)
[1] 0.01915675
```

The value we calculated is 0.019, which is smaller than 0.05, so it is not a adequate model at 95%, so we reject that given the defendant's race, sentence is independent of the victim's race.

For (a).iv, we create model.c4 for this part, here is the code:

```
> #(a).iv
> model.c4 <- glm(value ~ VR * DP + VR * DR, family=poisson)
> summary(model.c4)

Call:
glm(formula = value ~ VR * DP + VR * DR, family = poisson)

Deviance Residuals:
    1      2      3      4      5      6      7      8 
-0.47967  0.18976 -0.98198  0.16368  0.70243 -0.29660  0.20237 -0.04887 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   4.5797      0.1011  45.314 < 2e-16 ***
VR             -0.5876      0.1639  -3.586 0.000336 ***
DP             -2.8717      0.4196  -6.843 7.75e-12 ***
DR             -2.4375      0.3476  -7.013 2.34e-12 ***
VR:DP          1.0579      0.4635   2.282 0.022471 *
VR:DR          3.3116      0.3786   8.748 < 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: 395.9153  on 7  degrees of freedom
Residual deviance:   1.8819  on 2  degrees of freedom
AIC: 49.563
```

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(1.8819, df=2)
[1] 0.3902569
```

The value we calculated is 0.390, which is much greater than 0.05, so it is not a adequate model at 95%, so we accept that given the victim's race, sentence is independent of the defendant's race.

### 3.(b)

Three parts of (a) can be solved by using logistic regression models, they are (a).ii, (a).iii and (a).iv

First we preprocess the data, here is the code:

```

> #(b)
> DR.1 <- c(1,1,0,0)
> VR.1 <- c(1,0,1,0)
> DP_y.1 <- c(19,0,11,6)
> DP_n.1 <- c(132,9,52,97)
> total <- DP_n.1 + DP_y.1

```

Then we consider (a).ii, we create model.c5 for this part, here is the code:

```

> #(b) test (a).ii
> model.c5 <- glm(DP_y.1/total ~ 1, family=binomial, weights=total)
> summary(model.c5)

Call:
glm(formula = DP_y.1/total ~ 1, family = binomial, weights = total)

Deviance Residuals:
    1      2      3      4 
0.5921 -1.4513  1.5140 -1.8392

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0864      0.1767  -11.81  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.1316  on 3  degrees of freedom
Residual deviance: 8.1316  on 3  degrees of freedom
AIC: 22.445

```

Then we calculate the p-value for the deviance residual:

```

> 1 - pchisq(8.1316, df=3)
[1] 0.0433688

```

The value we calculated is 0.043, which is smaller than 0.05, so it is not a adequate model, so we reject that sentence is independent of both the defendant's and the victim's race.

Then we consider (a).iii, we create model.c6 for this part, here is the code:

```

> #(b) test (a).iii
> model.c6 <- glm(DP_y.1/total ~ DR.1, family=binomial, weights=total)
> summary(model.c6)

Call:
glm(formula = DP_y.1/total ~ DR.1, family = binomial, weights = total)

Deviance Residuals:
    1      2      3      4 
0.2666 -1.5085  1.7371 -1.5957

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1707      0.2560  -8.48  <2e-16 ***
DR.1          0.1664      0.3539   0.47   0.638
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.1316  on 3  degrees of freedom
Residual deviance: 7.9102  on 2  degrees of freedom
AIC: 24.224

```

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(7.9102, df=2)
[1] 0.01915675
```

The value we calculated is 0.019, which is smaller than 0.05, so it is not a adequate model, so we reject that given the defendant's race, sentence is independent of the victim's race.

Then we consider (a).iv, we create model.c7 for this part, here is the code:

```
> #(b) test (a).iv
> model.c7 <- glm(DP_y.1/total ~ VR.1, family=binomial, weights=total)
> summary(model.c7)
```

```
Call:
glm(formula = DP_y.1/total ~ VR.1, family = binomial, weights = total)
```

```
Deviance Residuals:
    1      2      3      4
-0.5158 -0.9955  0.7625  0.2082
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.8717      0.4196  -6.843 7.75e-12 ***
VR.1           1.0579      0.4635   2.282  0.0225 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 8.1316 on 3 degrees of freedom
Residual deviance: 1.8819 on 2 degrees of freedom
AIC: 18.196
```

Then we calculate the p-value for the deviance residual:

```
> 1 - pchisq(1.8819, df=2)
[1] 0.3902569
```

The value we calculated is 0.390, which is larger than 0.05, so it is an adequate model, it means we can accept that given the victim's race, sentence is independent of the defendant's race.