**Problem 1**

Before beginning this problem, score values were assigned for the levels of each variable:

* Social Status (*S*): {1 = Lower, 2 = Lower Middle, 3 = Upper Middle, 4 = Higher}
* Parental Encouragement (*E*): {1 = Low, 2 = High}
* College Plans (*P*): {1 = No, 2 = Yes}

In *R*, I stored the data set with a data.frame structure, using the variable counts to store frequency counts, S as the level of social status, E as the level of parental encouragement, and P as whether or not Minnesota male high school seniors plan to attend college:

> ####Data Storage

> counts<-c(749,627,420,153,233,330,374,266,35,38,37,26,133,303,467,800)

> S<-rep(c("Lower","Lower Middle","Upper Middle","Higher"),4)

> E<-rep(c(rep("Low",4),rep("High",4)),2)

> P<-c(rep("No",8),rep("Yes",8))

> data<-data.frame(S,E,P,counts)

Data were displayed as a contingency table to ensure the values were input correctly:

> ####Contingency Table

> cont.table<-xtabs(formula=counts~S+E+P,data=data);cont.table

, , P = No

E

S High Low

Higher 266 153

Lower 233 749

Lower Middle 330 627

Upper Middle 374 420

, , P = Yes

E

S High Low

Higher 800 26

Lower 133 35

Lower Middle 303 38

Upper Middle 467 37

1. Using the model (*SE*, *EP*, *SP*), find the conditional odds ratios between *E* and *S*. Construct a 95% confidence interval for the conditional odds ratio between *P* and *E*. Interpret.

To solve this problem, I first fit the data to the loglinear pairwise dependent model (*SE*, *EP*, *SP*) referred to as *M*1. Note that the values of for the highest factor levels in each main effect and two-way interaction were set to 0 in *R*. For instance, in the case of *S*, this approach is analogous to using 3 dummy variables:

* *S*1 = {1: Lower, 0: otherwise}
* *S*2 = {1: Lower Middle, 0: otherwise}
* *S*3 = {1: Upper Middle, 0: otherwise}

*M*1 can thus be expressed as:

In *R*, *M*1 was stored in fit:

> ###(SE,SP,PE) model

> fit<-glm(formula=counts~.^2,data=data,family=poisson(link="log"),na.action=na.exclude,control=list(epsilon=0.0001,maxit=50,trace=T))

Deviance = 1.578687 Iterations - 1

Deviance = 1.575468 Iterations - 2

Deviance = 1.575468 Iterations - 3

> summary(fit)

Call:

glm(formula = counts ~ .^2, family = poisson(link = "log"), data = data, na.action = na.exclude, control = list(epsilon = 1e-04, maxit = 50, trace = T))

Deviance Residuals:

1 2 3 4 5 6 7 8

-0.15119 0.04135 -0.04446 0.32807 0.27320 -0.05691 0.04719 -0.24539

9 10 11 12 13 14 15 16

0.73044 -0.16639 0.15116 -0.75147 -0.35578 0.05952 -0.04217 0.14245

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 5.59850 0.05886 95.116 < 2e-16 \*\*\*

SLower -0.16542 0.08573 -1.930 0.05366 .

SLower Middle 0.20372 0.07841 2.598 0.00937 \*\*

SUpper Middle 0.32331 0.07664 4.219 2.46e-05 \*\*\*

ELow -0.59471 0.09234 -6.441 1.19e-10 \*\*\*

PYes 1.08107 0.06731 16.060 < 2e-16 \*\*\*

SLower:ELow 1.78588 0.11444 15.606 < 2e-16 \*\*\*

SLower Middle:ELow 1.23178 0.10987 11.211 < 2e-16 \*\*\*

SUpper Middle:ELow 0.71532 0.11136 6.424 1.33e-10 \*\*\*

SLower:PYes -1.59311 0.11527 -13.820 < 2e-16 \*\*\*

SLower Middle:PYes -1.17298 0.09803 -11.965 < 2e-16 \*\*\*

SUpper Middle:PYes -0.85460 0.09259 -9.230 < 2e-16 \*\*\*

ELow:PYes -2.68292 0.09867 -27.191 < 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: 3211.0014 on 15 degrees of freedom

Residual deviance: 1.5755 on 3 degrees of freedom

AIC: 141.39

Number of Fisher Scoring iterations: 3

Next, I stored the coefficients for two-way interactions between *S* and *E* as variables in the format lmdSE.ik, where i is the level of *S*, and k is the level of *E*. Likewise, I stored the coefficients for two-way interactions between *E* and *P* as variables in the format lmdPE.jk, where j is the level of *P*, and k is the level of *E*:

> ###Lambda coefficients

> lmdSE.11<-summary(fit)$coef[7]

> lmdSE.12<-0

> lmdSE.21<-summary(fit)$coef[8]

> lmdSE.22<-0

> lmdSE.31<-summary(fit)$coef[9]

> lmdSE.32<-0

> lmdSE.41<-0

> lmdSE.42<-0

> lmdPE.11<-summary(fit)$coef[13]

> lmdPE.12<-0

> lmdPE.21<-0

> lmdPE.22<-0

Since Social Status (*S*) has 4 levels, six partial odds ratios are computed to compare the odds of Low Parental Encouragement (*E*) on two levels of *S* without dependence on *P*. The computational methods are displayed in Table 1, along with variable names used for in *R*.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1**: Computation of conditional odds ratios for Social Status (*S*) and Parental Encouragement (*E*). | | | |
| *S* | *E* |  | in *R* |
| Lower  Lower Middle | Low |  | ORSE.1 |
| Lower Middle  Upper Middle | Low |  | ORSE.2 |
| Upper Middle  Higher | Low |  | ORSE.3 |
| Lower  Higher | Low |  | ORSE.4 |
| Lower  Upper Middle | Low |  | ORSE.5 |
| Lower Middle  Higher | Low |  | ORSE.6 |

In *R*, each odds ratio was calculated separately by exponentiating each :

> ###SE Local ORs

> #S(Lower, Lower Middle)

> ORSE.1<-exp((lmdSE.11+lmdSE.22)-(lmdSE.12+lmdSE.21));ORSE.1

[1] 1.740375

> #S(Lower Middle, Upper Middle)

> ORSE.2<-exp((lmdSE.21+lmdSE.22)-(lmdSE.31+lmdSE.31));ORSE.2

[1] 0.8196692

> #S(Upper Middle, Higher)

> ORSE.3<-exp((lmdSE.31+lmdSE.32)-(lmdSE.42+lmdSE.41));ORSE.3

[1] 2.044833

> #S(Lower, Higher)

> ORSE.4<-exp((lmdSE.11+lmdSE.41)-(lmdSE.42+lmdSE.12));ORSE.4

[1] 5.964818

> #S(Lower, Upper Middle)

> ORSE.5<-exp((lmdSE.11+lmdSE.31)-(lmdSE.32+lmdSE.12));ORSE.5

[1] 12.19706

> #S(Lower Middle, Higher)

> ORSE.6<-exp((lmdSE.21+lmdSE.41)-(lmdSE.42+lmdSE.22));ORSE.6

[1] 3.427318

For clarity, the odds ratios and their inverse equivalents are displayed in Table 2, rounded to three decimal places.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 2**: Conditionala between Parental Encouragement (*E*) and Social Status (*S*) | | | | | |
|  | Social Status (*Sb*) | | | | |
| Social Status  (*Sa*) |  | Lower | Lower Middle | Upper Middle | Higher |
| Lower | - | 1.740 | 12.197 | 5.965 |
| Lower Middle | 0.575 | - | 0.820 | 3.427 |
| Upper Middle | 0.082 | 1.220 | - | 2.045 |
| Higher | 0.168 | 0.292 | 0.489 | - |

aFor each cell, odds ratios are calculated in the format .

Overall, the conditional odds of receiving low parental encouragement are lower for those in Higher social status than for any other group, while the conditional odds for those in the Lower social status are higher than for any other group.

Estimated partial conditional odd ratios are further interpreted in the context of Problem 1 as follows. Irrespective of whether Minnesota male high school seniors plan to go to college…

* there is a strong association between Social Status (*S*) and Parental Encouragement (*E*) in Lower vs. Upper Middle or Higher social status. The odds of receiving low parental encouragement for those in Lower social statuses are 12.197 times the odds for those in Upper Middle social statuses; likewise, the odds of receiving low parental encouragement for those in Lower social statuses are 5.965 times the odds for those in Higher social statuses.
* there is a moderate association between Social Status (*S*) and Parental Encouragement (*E*) in Higher vs. Lower Middle or Upper Middle social status. The odds of receiving low parental encouragement for those in Lower Middle social statuses are 3.427 times the odds for those in higher social statuses; likewise, the odds of receiving low parental encouragement for those in Upper Middle statuses are 2.045 times the odds for those in Higher social statuses.
* there is a weak association between Social Status (*S*) and Parental Encouragement (*E*) in Lower vs. Lower Middle and Upper Middle vs. Lower Middle social status. The odds of receiving low parental encouragement for those in Lower social statuses are 1.740 times the odds for those in Lower Middle social statuses; likewise, the odds of receiving low parental encouragement for those in Upper middle statuses are 1.220 times the odds for those in Lower Middle social statuses.
* there is a trend suggesting that those in Higher social statuses often have Lower odds of receiving low parental encouragement. However, there are two exceptions to this trend:
  + The Lower-Upper Middle conditional odds ratio (12.197) is larger than the Lower-Higher conditional odds ratio (5.965).
  + The odds of receiving low parental encouragement for those in Upper Middle statuses are 1.220 times the odds for those in Lower Middle social statuses.

For the conditional odds ratio between *P* and *E*, the following formula was used:

In addition to calculating the estimated odds ratio for *P* and *E*, a corresponding 95% confidence interval was calculated with the use of the following formula:

In *R*, the method was applied using the lambda coefficients stored earlier:

> ###PE OR

> ln.ORPE<-(lmdPE.11+lmdPE.22)-(lmdPE.12+lmdPE.21)

> ORPE<-exp(ln.ORPE);ORPE

[1] 0.06836293

As such, the estimated odds ratio is = 0.068. Moreover, a 95% confidence interval was first calculated for in *R*. The interval bounds were then exponentiated separately to obtain bounds for :

> ###PE OR + 95% CI

> ln.ORPE<-(lmdPE.11+lmdPE.22)-(lmdPE.12+lmdPE.21)

> ORPE<-exp(ln.ORPE);ORPE

[1] 0.06836293

> z\_alpha=qnorm(0.025,lower.tail=F)

> ln.ORPE.se<-coef(summary(fit))[13, "Std. Error"]

> ORPE.ci<-exp(ln.ORPE+z\_alpha\*c(-1,1)\*ln.ORPE.se);ORPE.ci

[1] 0.05634226 0.08294822

From this output, we can ascertain that = [0.056, 0.083]. Thus, with 95% confidence, the odds ratio for *E* and *P* is between 0.056 and 0.083. This confidence interval implies that given a level of Social Status (*S*), Parental Encouragement (*E*) and College Plans (*P*) are conditionally associated.

In order to better understand this association within the context of Problem 1, it is helpful to consider the inverse equivalent of the estimated odds ratio—14.628. This value indicates that given a level of Social Status (*S*), the odds of planning to go to college for those who receive High parental encouragement are 14.628 times the odds for those who receive Low parental encouragement.

1. Check the goodness-of fit and residuals of this model.

The goodness-of-fit for *M*1 was tested using the following hypothesis test:

* *H*0: The (*SE*, *EP*, *SP*) model is an adequate fit for the data set.
* *Ha*: The (*SE*, *EP*, *SP*) model is not an adequate fit for the data set.

The likelihood ratio test (LRT) statistic, *G*2, was obtained from the residual deviance of the model *M*1. Assuming *H*0 is true, *G*2 follows a χ2 distribution; thus, a corresponding *p*-value was also calculated. In *R*:

> ###Goodness-of-fit

> G.sq<-fit$deviance;G.sq

[1] 1.575468

> G.sq.df<-fit$df.residual;G.sq.df

[1] 3

> p.val<-round(1-pchisq(G.sq,G.sq.df),4);p.val

[1] 0.665

The likelihood ratio test (LRT) indicates *G*2(3) = 1.575 is not significant (*p* = .665). Hence, *H*0 is not rejected; the loglinear model *M*1 fits the data adequately.

Standardized Pearson residuals for the model were also analyzed with *R*. A contingency table was used to display the residuals:

> ###Residuals

> save.predict<-predict(object=fit,type="response")

> save.pearson<-residuals(object=fit,type="pearson")

> h<-lm.influence(model=fit)$h

> standard.pearson<-save.pearson/sqrt(1-h)

> save.all<-data.frame(data,predict=round(save.predict,4),pearson=round(save.pearson,4),standard.pearson=round(standard.pearson,4))

> xtabs(standard.pearson~S+E+P,data=save.all)

, , P = No

E

S High Low

Higher -0.9637 0.9637

Lower 0.9965 -0.9965

Lower Middle -0.2250 0.2250

Upper Middle 0.2006 -0.2006

, , P = Yes

E

S High Low

Higher 0.9637 -0.9637

Lower -0.9965 0.9965

Lower Middle 0.2250 -0.2250

Upper Middle -0.2006 0.2006

From this output, we can observe that there are absolutely no outliers; in fact, every standardized residual lies within one standard deviation of the mean. This fact provides evidence that the loglinear model *M*1 fits the data set well and makes adequate predictions of , since the variance in standardized residuals seems to be constant.

1. Given the model (*SE*, *EP*, *SP*), test whether *E* and *P* are conditionally-independent. Interpret.

To test whether *E* and *P* are conditionally independent, a reduced loglinear model, *M*2 was defined as the loglinear model (*SE*, *SP*). This model is similar to the complex model *M*1, except that it lacks the interaction term . Thus, *M*2 assumes that *E* and *P* are conditionally independent given *S*. In *R*, *M*2 was stored in fit2:

> ###(SE,SP) model

> fit2<-glm(counts~(E+S)^2+(P+S)^2,data=data,family=poisson(link="log"),na.action=na.exclude,control=list(epsilon=0.0001,maxit=50,trace=T))

Deviance = 1428.552 Iterations - 1

Deviance = 1097.683 Iterations - 2

Deviance = 1083.878 Iterations - 3

Deviance = 1083.827 Iterations - 4

> summary(fit2)

Call:

glm(formula = counts ~ (E + S)^2 + (P + S)^2, family = poisson(link = "log"),

data = data, na.action = na.exclude, control = list(epsilon = 1e-04,

maxit = 50, trace = T))

Deviance Residuals:

1 2 3 4 5 6 7 8

3.016 5.916 7.813 9.984 -4.713 -6.682 -6.512 -5.135

9 10 11 12 13 14 15 16

-8.723 -12.549 -12.841 -10.322 9.129 9.496 7.296 3.416

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 5.88265 0.05019 117.213 < 2e-16 \*\*\*

ELow -1.78419 0.08029 -22.221 < 2e-16 \*\*\*

SLower -0.13794 0.07347 -1.877 0.0605 .

SLower Middle 0.26305 0.06613 3.978 6.95e-05 \*\*\*

SUpper Middle 0.36045 0.06478 5.564 2.64e-08 \*\*\*

PYes 0.67872 0.05992 11.327 < 2e-16 \*\*\*

ELow:SLower 2.54596 0.10223 24.903 < 2e-16 \*\*\*

ELow:SLower Middle 1.83350 0.09761 18.785 < 2e-16 \*\*\*

ELow:SUpper Middle 1.17428 0.09909 11.850 < 2e-16 \*\*\*

SLower:PYes -2.44434 0.10260 -23.824 < 2e-16 \*\*\*

SLower Middle:PYes -1.71064 0.08692 -19.681 < 2e-16 \*\*\*

SUpper Middle:PYes -1.13323 0.08264 -13.712 < 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: 3211.0 on 15 degrees of freedom

Residual deviance: 1083.8 on 4 degrees of freedom

AIC: 1221.6

Number of Fisher Scoring iterations: 4

To test whether *P* and *E* are conditionally independent given *S*, I compared the residual deviance of *M*1 to that of *M*2. Since the only difference between the models is the presence of the interaction term between *P* and *E*, if the difference in residual deviance between the models is small, it may be possible to remove the interaction term—thus indicating that *P* and *E* are conditionally independent given *S*. Therefore, I conduct the following hypothesis test:

* *H*0: .
* *Ha*: At least one .

In *R*, the test statistic for this hypothesis test was calculated as :

> ###Test if P and E are conditionally independent

> G.sq.S<-fit2$deviance-fit$deviance;G.sq.S

[1] 1082.251

> G.sq.df.S<-fit2$df.residual- fit$df.residual;G.sq.df.S

[1] 1

> p.val.S<-round(1-pchisq(G.sq.S,G.sq.df.S),4);p.val.S

[1] 0

The test statistic (1) = 1082.251 is significant (*p* < .001). Therefore, there is evidence to suggest that College Plans (*P*) and Parental Encouragement (*E*) are conditionally associated given Social Status (*S*). Hence, we reject simplifying to the reduced model *M*2; further analyses must continue to use the complex model *M*1.

1. Treating whether or not they have plans to attend college, (*P*), as the response variable, obtain an equivalent logit model to the loglinear model (*SE*, *EP*, *SP*). Interpret the effects on the response. You should use the fit of the loglinear model to obtain the logit model as done in class.

Recall that the pairwise dependent loglinear model *M*1 is given by the following:

To obtain a logit model using College Plans (*P*) as the response variable, I took the natural log of the odds ratio for the two levels of *P*:

In this expression, the values for are as follows:

Therefore, simplifying the logit model:

In *M*1, values are constrained to 0. Thus, the expression can be simplified further:

At this juncture, I defined a logit model as *M*3, with *P* as the response variable:

, where , , and .

Without fitting *M*3 to the data set, we can observe that the logit of *P* derived from *M*1 is dependent upon various elements from *M*1:

* : the main effect of *P*, which serves as the intercept in *M*3. For all Minnesota male high school seniors, this multiplies the odds of *P* by .
* : the two-way interaction term between *S* and *P*. For a given level of Social Status (*S*), the estimated odds of *P* are multiplied by .
* : the two-way interaction term between *P* and *E*. For a given level of Parental Encouragement (*E*), the estimated odds of *P* are multiplied by .

**Problem 2**

Before beginning this problem, score values were assigned for the levels of each variable:

* Victims’ Race: {0 = Black, 1 = White }
* Defendant’s Race: {0 = Black, 1 = White }
* Death Penalty: {1 = Yes, 2 = No}

In *R*, I stored the data set with a data.frame structure, using the variable pct to store the proportions of death penalty verdicts, Victim as the Victims’ Race, and Defendant as the Defendant’s Race.

> ####Data Storage

> pct<-c(.113,.229,0,.028)

> Victim<-c(rep("White",2),rep("Black",2))

> Defendant<-rep(c("White","Black"),2)

> data.2<-data.frame(Victim,Defendant,pct)

The data were then displayed in a contingency table:

> ####Contingency Table

> cont.table.2<-xtabs(formula=pct~Victim+Defendant,data=data.2);cont.table.2

Defendant

Victim Black White

Black 0.028 0.000

White 0.229 0.113

Lastly, a logit model was used to fit the data. As stated, the coefficients for the second level of each variable were coded as 0.

> ###Logit model

> fit3<-glm(formula=pct~Victim+Defendant,data=data.2,family=binomial(logit))

Warning message:

In eval(expr, envir, enclos) : non-integer #successes in a binomial glm!

> summary(fit3)

Call:

glm(formula = pct ~ Victim + Defendant, family = binomial(logit),

data = data.2)

Deviance Residuals:

1 2 3 4

0.02505 -0.01834 -0.12489 0.05217

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -3.8800 6.1225 -0.634 0.526

VictimWhite 2.7096 6.3261 0.428 0.668

DefendantWhite -0.9699 3.8503 -0.252 0.801

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 0.429219 on 3 degrees of freedom

Residual deviance: 0.019283 on 1 degrees of freedom

AIC: 6.8192

Number of Fisher Scoring iterations: 7

The model can thus be interpreted as:

1. Interpret the parameter estimates. Which group is most likely to have the yes response? Find the estimated probability in that case.

The model can be rewritten to compare the odds of a death penalty verdict:

Higher odds allows us to identify the group most likely to be given the death penalty. Consider the estimates for these parameters:

* = 2.710: The odds of being given the death penalty for a White victim are times the odds for a Black victim, given the defendant’s race.
* = -0.970: The odds of being given the death penalty for a White defendant are times the odds for a Black defendant, given the victim’s race.

From this analysis, we can expect that the group most likely to be sentenced to death is the one with a Black defendant and a White victim. To find the e probability, we can modify the formula used for since :

Hence for a Black defendant and a White victim:

= 0.237.

1. Interpret the 95% confidence interval for the conditional odds ratio between death penalty and defendant’s race.

By definition, for logit models…

I adapted this formula to estimate the conditional odds ratio between death penalty and defendant’s race:

= 0.379.

In *R*, I used the estimated value of to obtain a 95% confidence interval for . Exponentiating both the upper and lower bounds, I obtained a 95% confidence interval for :

> ###Conditional OR between defendant race and death penalty

> ln.ORDefVer<-summary(fit3)$coef[3]

> ORDefVer<-exp(ln.ORDefVer);ORDefVer

[1] 0.3791157

> ln.ORDefVer.se<-coef(summary(fit3))[3, "Std. Error"]

> ORDefVer.ci<-exp(ln.ORDefVer+z\_alpha\*c(-1,1)\*ln.ORDefVer.se);ORDefVer.ci

[1] 2.001517e-04 7.180986e+02

From the output, we can observe that = [2.002 × 10-4, 718.099]. With 95% confidence, given either a black or white victim, the odds of a White defendant receiving the death penalty are between 2.002 × 10-4 and 718.099 times the odds of a Black defendant receiving the death penalty. Because this range includes 1, we cannot reject the null hypothesis that verdict and defendant’s race are conditionally independent given a victim’s race.

1. Test the effect of defendant’s race, controlling for victim’s race, using… Interpret.
   1. Wald test

Testing the effect of defendant’s race requires the use of the following hypothesis test:

* *H*0: = 0.
* *Ha*: = 0.

From the summary of the logit fit for this data set, we can obtain the statistics necessary to perform the Wald Test:

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -3.8800 6.1225 -0.634 0.526

VictimWhite 2.7096 6.3261 0.428 0.668

DefendantWhite -0.9699 3.8503 -0.252 0.801

The test statistic for the parameter is *z* = -0.252, which is not significant (*p* = .801). Thus, there is insufficient evidence to indicate that there is an association between death penalty verdict and defendant’s race. In other words, for both Black and White victims the odds of a White defendant being given the death penalty are not significantly different from the odds of a Black defendant being given the death penalty.

* 1. Likelihood ratio test

For this problem, I first fit the data set in *R* into a reduced model not accounting for the defendant’s race:

> ###Reduced logit model

> fit4<-glm(formula=pct~Victim,data=data.2,family=binomial(logit))

Warning message:

In eval(expr, envir, enclos) : non-integer #successes in a binomial glm!

> summary(fit4)

Call:

glm(formula = pct ~ Victim, family = binomial(logit), data = data.2)

Deviance Residuals:

1 2 3 4

-0.1623 0.1480 -0.1679 0.1050

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -4.255 6.018 -0.707 0.480

VictimWhite 2.676 6.305 0.424 0.671

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 0.429219 on 3 degrees of freedom

Residual deviance: 0.087473 on 2 degrees of freedom

AIC: 4.8065

Number of Fisher Scoring iterations: 7

By comparing the residual deviance of this reduced model to the complex, we can investigate the effect of the defendants’ race in isolation since victims’ race is a coefficient in both models. Conditional independence was tested with the following hypothesis test:

* *H*0: .
* *Ha*: .

The likelihood ratio test (LRT) statistic, *G*2, was obtained from the residual deviance of the complex logit model. Assuming *H*0 is true, *G*2 follows a χ2 distribution; thus, a corresponding *p*-value was also calculated. In *R*:

> ###Test effect of defendant’s race controlling for victim’s race

> G.sq.defendant<-fit4$deviance-fit3$deviance;G.sq.defendant

[1] 0.06819013

> G.sq.defendant.df<-fit4$df.residual- fit3$df.residual;G.sq.defendant.df

[1] 1

> p.val.defendant<-round(1-pchisq(G.sq.defendant,G.sq.defendant.df),4);p.val.defendant

[1] 0.794

The likelihood ratio test (LRT) indicates *G*2(1) = 0.068 is not significant (*p* = .794). Hence, *H*0 is not rejected; there is insufficient evidence to indicate that there is an association between verdict and defendant’s race. In other words, for both Black and White victims the odds of a White defendant being given the death penalty are not significantly different from the odds of a Black defendant being given the death penalty.