> ######Exam II

> #####Problem 1

> ####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)

>

> ####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

>

> ####(a)

> ###(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

>

> ###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

>

> ###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

>

> ###PE OR

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

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

[1] 0.06836293

>

> ###PE OR 95% CI

> 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

>

> ####(b)

> ###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

>

> ###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

>

> ####(c)

> ###(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

> ###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

>

> #####Problem 2

> ####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)

>

> ####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

>

> ####(a)

> ###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

>

> ####(b)

> ###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

>

>

> ####(c)

> ###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

>

> ###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