Chad Huntebrinker Homework 12

Chad Huntebrinker

2025-04-28

# Problem 7.3a

We have table that record automobile accidents and whether the person had a seatbelt on or not. We need to find a loglinear model that describes the data well and interpret the associations. In the model below, ‘S’ stands for whether the seatbelt worn or not, ‘E’ is for whether the person was ejected or not, and ‘I’ is for whether the injury was fatal or nonfatal.

##   
## Call:  
## glm(formula = count ~ S \* E \* I, family = poisson, data = motor\_data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.91572 0.03150 219.567 <2e-16 \*\*\*  
## Sused -0.73571 0.05534 -13.294 <2e-16 \*\*\*  
## Eyes -0.70713 0.05481 -12.902 <2e-16 \*\*\*  
## Inonfatal 5.05045 0.03160 159.836 <2e-16 \*\*\*  
## Sused:Eyes -2.83383 0.27659 -10.246 <2e-16 \*\*\*  
## Sused:Inonfatal 1.69615 0.05542 30.606 <2e-16 \*\*\*  
## Eyes:Inonfatal -2.82003 0.05680 -49.644 <2e-16 \*\*\*  
## Sused:Eyes:Inonfatal 0.44197 0.27863 1.586 0.113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 1.6249e+06 on 7 degrees of freedom  
## Residual deviance: 2.8231e-11 on 0 degrees of freedom  
## AIC: 92.999  
##   
## Number of Fisher Scoring iterations: 3

We see the p-value for the full model’s 3-way interaction term is 0.113. Since it’s > 0.05, we should try removing it from the model and see what we get.

##   
## Call:  
## glm(formula = count ~ S + E + I + S:E + S:I + E:I, family = poisson,   
## data = motor\_data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.92251 0.03110 222.56 <2e-16 \*\*\*  
## Sused -0.75682 0.05394 -14.03 <2e-16 \*\*\*  
## Eyes -0.72784 0.05345 -13.62 <2e-16 \*\*\*  
## Inonfatal 5.04362 0.03120 161.65 <2e-16 \*\*\*  
## Sused:Eyes -2.39964 0.03334 -71.97 <2e-16 \*\*\*  
## Sused:Inonfatal 1.71732 0.05402 31.79 <2e-16 \*\*\*  
## Eyes:Inonfatal -2.79779 0.05526 -50.63 <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: 1.6249e+06 on 7 degrees of freedom  
## Residual deviance: 2.8540e+00 on 1 degrees of freedom  
## AIC: 93.853  
##   
## Number of Fisher Scoring iterations: 3

What we find is AIC is just as good for model\_2 as it is model\_1 and the deviance for model\_2 is small too (it’s 2.854). And since model\_2 is a simpler model, we’ll stick with that one. We see that the log-odds for S:E is e^(-2.399) which is about 0.091, SI is e^(1.717) = 5.567, and E:I is e^(-2.797) = 0.061. That means that seatbelts reduce ejection odds by 91%, seatbelts increase the odds of surviving an accident by 5.567 times, and ejection reduces the chances of surviving an accident by 94%.

# Problem 8.1

We need to apply the McNemar test to the following table and interpret. Our null hypothesis is that there is no association between smoking status and birth weight. Our alternative hypothesis is that smoking is associated with birth weight status, specifically low birth weights are more associated with a smoking status in the data.

## LBW  
## NBW nonsmoke smoke  
## nonsmoke 159 22  
## smoke 8 14

##   
## McNemar's Chi-squared test  
##   
## data: tab  
## McNemar's chi-squared = 6.5333, df = 1, p-value = 0.01059

We see that the McNemar’s chi-squared statistic is 6.53 and the p-value is about 0.011. Since that is < 0.05, we reject the null hypothesis and have strong evidence that low birth weight cases are more likely than normal birth weights to be smokers.

#Code Appendix  
#Chad Huntebrinker  
#Problem 7.3  
  
#Part a  
#We have table that record automobile accidents and whether the person had a seatbelt on or not  
#We need to find a loglinear model that describes the data well and interpret the associations.  
  
motor\_data <- data.frame(  
 S = c("used", "used", "used", "used", "not used", "not used", "not used", "not used"),  
 E = c("yes", "yes", "no", "no", "yes", "yes", "no", "no"),  
 I = c("nonfatal", "fatal", "nonfatal", "fatal", "nonfatal", "fatal", "nonfatal", "fatal"),  
 count = c(1105, 14, 411111, 483, 4624, 497, 157342, 1008)  
)  
  
model\_1 <- glm(count ~ S\*E\*I, family = poisson, data = motor\_data)  
summary(model\_1)

##   
## Call:  
## glm(formula = count ~ S \* E \* I, family = poisson, data = motor\_data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.91572 0.03150 219.567 <2e-16 \*\*\*  
## Sused -0.73571 0.05534 -13.294 <2e-16 \*\*\*  
## Eyes -0.70713 0.05481 -12.902 <2e-16 \*\*\*  
## Inonfatal 5.05045 0.03160 159.836 <2e-16 \*\*\*  
## Sused:Eyes -2.83383 0.27659 -10.246 <2e-16 \*\*\*  
## Sused:Inonfatal 1.69615 0.05542 30.606 <2e-16 \*\*\*  
## Eyes:Inonfatal -2.82003 0.05680 -49.644 <2e-16 \*\*\*  
## Sused:Eyes:Inonfatal 0.44197 0.27863 1.586 0.113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 1.6249e+06 on 7 degrees of freedom  
## Residual deviance: 2.8231e-11 on 0 degrees of freedom  
## AIC: 92.999  
##   
## Number of Fisher Scoring iterations: 3

#We see the p-value for the full model's 3-way interaction term is 0.113. Since it's > 0.05, we should  
#try removing it from the model and see what we get.  
  
model\_2 <- glm(count ~ S+E+I+S:E+S:I+E:I, family = poisson, data = motor\_data)  
summary(model\_2)

##   
## Call:  
## glm(formula = count ~ S + E + I + S:E + S:I + E:I, family = poisson,   
## data = motor\_data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.92251 0.03110 222.56 <2e-16 \*\*\*  
## Sused -0.75682 0.05394 -14.03 <2e-16 \*\*\*  
## Eyes -0.72784 0.05345 -13.62 <2e-16 \*\*\*  
## Inonfatal 5.04362 0.03120 161.65 <2e-16 \*\*\*  
## Sused:Eyes -2.39964 0.03334 -71.97 <2e-16 \*\*\*  
## Sused:Inonfatal 1.71732 0.05402 31.79 <2e-16 \*\*\*  
## Eyes:Inonfatal -2.79779 0.05526 -50.63 <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: 1.6249e+06 on 7 degrees of freedom  
## Residual deviance: 2.8540e+00 on 1 degrees of freedom  
## AIC: 93.853  
##   
## Number of Fisher Scoring iterations: 3

#What we find is AIC is just as good for model\_2 as it is model\_1 and the deviance for model\_2 is small  
#too (it's 2.854). And since model\_2 is a simpler model, we'll stick with that one.  
#We see that the log-odds for S:E is e^(-2.399) which is about 0.091, SI is e^(1.717) = 5.567,  
#and E:I is e^(-2.797) = 0.061. That means that seatbelts reduce ejection odds by 91%, seatbelts increase the  
#odds of surviving an accident by 5.567 times, and ejection reduces the chances of surviving an accident by  
#94%.  
  
#Chad Huntebrinker  
#Problem 8.1  
  
#We need to apply the McNemar test to the following table and interpret. Our null hypothesis is that  
#there is no association between smoking status and birth weight. Our alternative hypothesis is that  
#smoking is associated with birth weight status, specifically low birth weights are more associated  
#with a smoking status in the data.  
  
smoker\_data <- data.frame(  
 NBW = c("nonsmoke", "nonsmoke", "smoke", "smoke"),  
 LBW = c("nonsmoke", "smoke","smoke","nonsmoke"),  
 count = c(159, 22, 14, 8)  
)  
  
tab <- xtabs(count~NBW + LBW, data = smoker\_data)  
  
tab

## LBW  
## NBW nonsmoke smoke  
## nonsmoke 159 22  
## smoke 8 14

mcnemar.test(tab, correct = FALSE)

##   
## McNemar's Chi-squared test  
##   
## data: tab  
## McNemar's chi-squared = 6.5333, df = 1, p-value = 0.01059

#We see that the McNemar's chi-squared statistic is 6.53 and the p-value is about 0.011.  
#Since that is < 0.05, we reject the null hypothesis and have strong evidence that low birth  
#weight cases are more likely than normal birth weights to be smokers.