Chad Huntebrinker Homework 11

Chad Huntebrinker

2025-04-21

# Problem 6.2

For this problem, we need to use a model fit to estimate an odds ratio that describes the effect of length on primary food choice being either invertebrate or other. We will take the data and subset it so that it’s just invertebrate or other in our data.

##   
## Call:  
## vglm(formula = y ~ x, family = cumulative(parallel = TRUE, link = "logitlink"),   
## data = subset\_data)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.133 1.875 2.738 0.00619 \*\*  
## x -2.179 0.955 NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Name of linear predictor: logitlink(P[Y<=1])   
##   
## Residual deviance: 25.1233 on 26 degrees of freedom  
##   
## Log-likelihood: -12.5617 on 26 degrees of freedom  
##   
## Number of Fisher scoring iterations: 4   
##   
## Warning: Hauck-Donner effect detected in the following estimate(s):  
## 'x'  
##   
##   
## Exponentiated coefficients:  
## x   
## 0.1131819

## Odds ratio: 0.1131819

For each one meter increase in length, the odds that the alligator chooses invertebrate as its primary food decrease by about 89% (since e^(-2.179) = 0.113 and 1 - 0.113 = 0.887).

# Problem 6.8

We need to fit a cumulative logit model with a proportional odds structure, interpret the estimated treatment effect, and check whether a model allowing interaction provides a significant better fit.

##   
## Call:  
## vglm(formula = Response ~ Treatment + Gender, family = cumulative(parallel = TRUE,   
## link = "logitlink"), data = data, weights = Count)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 0.7526 0.4122 1.826 0.0679 .   
## (Intercept):2 1.9648 0.4304 4.565 4.99e-06 \*\*\*  
## TreatmentSequential -0.2152 0.2623 -0.820 0.4120   
## GenderMale -0.7149 0.4008 -1.784 0.0745 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])  
##   
## Residual deviance: 438.9876 on 20 degrees of freedom  
##   
## Log-likelihood: -219.4938 on 20 degrees of freedom  
##   
## Number of Fisher scoring iterations: 3   
##   
## No Hauck-Donner effect found in any of the estimates  
##   
##   
## Exponentiated coefficients:  
## TreatmentSequential GenderMale   
## 0.8064032 0.4892271

The logit model was fit to evaluate the effects of treatment type and gender on chemotherapy response. The estimated treatment effect comparing sequential to alternating therapy found the odds ratio of 0.81 (e^(-0.2152) = 0.81) with a p-value of 0.41. The estimated treatment effect comparing the two genders had an odds ratio of 0.49 (e^(-0.7149) = 0.49) with a p-value of 0.07. This showed that sequential treatment had about 19% lower odds of having a better remission (complete remission being the best type of remission) and males had a 51% lower odds of having a better remission. However, since both of these estimations had a p-value < 0.05, neither one of them are significant to the chemotherapy response.

Next, we’ll see if a model with interaction fits better. We’ll do a Likelihood ratio test to see if the interaction model gives us a better fit. Our null hypothesis is that the model with interaction does not provide a better fit, while our alternative hypothesis is that the model with the interaction does.

##   
## Call:  
## vglm(formula = Response ~ Treatment + Gender + Treatment:Gender,   
## family = cumulative(parallel = TRUE, link = "logitlink"),   
## data = data, weights = Count)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 0.63804 0.62120 1.027 0.30437   
## (Intercept):2 1.85060 0.63233 2.927 0.00343 \*\*  
## TreatmentSequential -0.03677 0.77797 -0.047 0.96231   
## GenderMale -0.58710 0.65449 -0.897 0.36970   
## TreatmentSequential:GenderMale -0.20233 0.82626 -0.245 0.80656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])  
##   
## Residual deviance: 438.9256 on 19 degrees of freedom  
##   
## Log-likelihood: -219.4628 on 19 degrees of freedom  
##   
## Number of Fisher scoring iterations: 3   
##   
## No Hauck-Donner effect found in any of the estimates  
##   
##   
## Exponentiated coefficients:  
## TreatmentSequential GenderMale   
## 0.9639003 0.5559371   
## TreatmentSequential:GenderMale   
## 0.8168274

## Likelihood ratio test  
##   
## Model 1: Response ~ Treatment + Gender  
## Model 2: Response ~ Treatment + Gender + Treatment:Gender  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 20 -219.49   
## 2 19 -219.46 -1 0.0621 0.8033

Our likelihood ratio test gives us a p-value of about 0.8. Since that is > 0.05, we can’t reject our null hypothesis. So in conclusion, our model with an interaction term does not give us a better fit than the simpler model.

#Code Appendix  
#Chad Huntebrinker  
#Problem 6.2  
library(readxl)  
library(VGAM)  
  
#We need to use a model fit to estimate an odds ratio that describes the effect of length on primary food  
#choice being either invertebrate or other.  
  
alligator\_data <- read\_excel("alligator\_data.xlsx")  
  
#We're subsetting the data to only keep I and O.  
subset\_data <- subset(alligator\_data, y %in% c("I", "O"))  
  
#Create the model  
model <- vglm(y ~ x, data = subset\_data, family = cumulative(parallel = TRUE, link = "logitlink"))  
summary(model)

##   
## Call:  
## vglm(formula = y ~ x, family = cumulative(parallel = TRUE, link = "logitlink"),   
## data = subset\_data)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.133 1.875 2.738 0.00619 \*\*  
## x -2.179 0.955 NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Name of linear predictor: logitlink(P[Y<=1])   
##   
## Residual deviance: 25.1233 on 26 degrees of freedom  
##   
## Log-likelihood: -12.5617 on 26 degrees of freedom  
##   
## Number of Fisher scoring iterations: 4   
##   
## Warning: Hauck-Donner effect detected in the following estimate(s):  
## 'x'  
##   
##   
## Exponentiated coefficients:  
## x   
## 0.1131819

#Now we get the odds ratio  
cat("Odds ratio:", exp(coef(model)["x"]), "\n")

## Odds ratio: 0.1131819

#For each one-unit increase in length, the odds that the alligator chooses Invertebrate (vs Other) as  
#its primary food decrease by about 89% (since 1 - 0.113 = 0.887).  
  
#Chad Huntebrinker  
#Problem 6.8  
#We need to fit a cumulative logit model with a proportional odds structure, interpret the estimated  
#treatment effect, and check whether a model allowing interaction provides a significant better fit.  
  
library(VGAM)  
library(readxl)  
  
#Create the dataframe  
data <- data.frame(  
 Treatment = factor(rep(c("Sequential", "Alternating"), each = 6)),  
 Gender = factor(rep(rep(c("Male", "Female"), each = 3), 2)),  
 Response = ordered(rep(c("NoChange", "Partial", "Complete"), 4),  
 levels = c("NoChange", "Partial", "Complete")),  
 Count = c(45, 29, 26,  
 12, 5, 2,  
 44, 20, 20,  
 7, 3, 1)  
)  
  
#Fit cumulative logit model  
model\_1 <- vglm(Response ~ Treatment + Gender,  
 family = cumulative(parallel = TRUE, link = "logitlink"),  
 weights = Count,  
 data = data)  
  
  
summary(model\_1)

##   
## Call:  
## vglm(formula = Response ~ Treatment + Gender, family = cumulative(parallel = TRUE,   
## link = "logitlink"), data = data, weights = Count)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 0.7526 0.4122 1.826 0.0679 .   
## (Intercept):2 1.9648 0.4304 4.565 4.99e-06 \*\*\*  
## TreatmentSequential -0.2152 0.2623 -0.820 0.4120   
## GenderMale -0.7149 0.4008 -1.784 0.0745 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])  
##   
## Residual deviance: 438.9876 on 20 degrees of freedom  
##   
## Log-likelihood: -219.4938 on 20 degrees of freedom  
##   
## Number of Fisher scoring iterations: 3   
##   
## No Hauck-Donner effect found in any of the estimates  
##   
##   
## Exponentiated coefficients:  
## TreatmentSequential GenderMale   
## 0.8064032 0.4892271

#The logit model was fit to evaluate the effects of treatment type and gender on chemotherapy response.  
#The estimated treatment effect comparing Sequential to Alternating therapy found the odds ratio  
#as = 0.81 (exp(-0.2152)) with a p-value of 0.41. The estimated treatment effect comparing  
#the two genders had an odds ratio of 0.49 (exp(-0.7149)) with a p-value of 0.07. This showed that  
#Sequential treatment had about 19% lower odds of having a better remission (complete remission  
#being the best) and 51% lower odds of females having a better remission. However, since both of these  
#estimations had a p-value < 0.05, neither one of them are significant to the response.  
  
#Next, we'll see if a model with interaction fits better.  
model\_2 <- vglm(Response ~ Treatment + Gender + Treatment:Gender,  
 family = cumulative(parallel = TRUE, link = "logitlink"),  
 weights = Count,  
 data = data)  
  
summary(model\_2)

##   
## Call:  
## vglm(formula = Response ~ Treatment + Gender + Treatment:Gender,   
## family = cumulative(parallel = TRUE, link = "logitlink"),   
## data = data, weights = Count)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 0.63804 0.62120 1.027 0.30437   
## (Intercept):2 1.85060 0.63233 2.927 0.00343 \*\*  
## TreatmentSequential -0.03677 0.77797 -0.047 0.96231   
## GenderMale -0.58710 0.65449 -0.897 0.36970   
## TreatmentSequential:GenderMale -0.20233 0.82626 -0.245 0.80656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])  
##   
## Residual deviance: 438.9256 on 19 degrees of freedom  
##   
## Log-likelihood: -219.4628 on 19 degrees of freedom  
##   
## Number of Fisher scoring iterations: 3   
##   
## No Hauck-Donner effect found in any of the estimates  
##   
##   
## Exponentiated coefficients:  
## TreatmentSequential GenderMale   
## 0.9639003 0.5559371   
## TreatmentSequential:GenderMale   
## 0.8168274

#We'll do a Likelihood ratio test to see if the interaction model gives us a better fit.  
#Our null hypothesis is that the model with interaction does not provide a better fit, while our  
#alternative hypothesis is that the model with the interaction does.  
lrtest(model\_1, model\_2)

## Likelihood ratio test  
##   
## Model 1: Response ~ Treatment + Gender  
## Model 2: Response ~ Treatment + Gender + Treatment:Gender  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 20 -219.49   
## 2 19 -219.46 -1 0.0621 0.8033

#With a p-value of 0.8, we can't reject our null hypothesis. So our model with an interaction term does not  
#give us a better fit.