STAC51\_A3

Rafsaan Sanvir

2024-04-02

1a)

data1 = data.frame(  
 X = c(5, 3, 4, 5, 4, 6, 5, 3, 6, 5),  
 Y = c(53, 34, 45, 49, 45, 77, 62, 37, 75, 54)  
)  
  
model1 = glm(formula = Y~X, data = data1, family = poisson(link = log))  
summary(model1)

##   
## Call:  
## glm(formula = Y ~ X, family = poisson(link = log), data = data1)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.78577 0.22159 12.572 < 2e-16 \*\*\*  
## X 0.25099 0.04481 5.601 2.13e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 35.1413 on 9 degrees of freedom  
## Residual deviance: 2.4824 on 8 degrees of freedom  
## AIC: 64.285  
##   
## Number of Fisher Scoring iterations: 3

estimate of model in written sheet

1b)

full\_model = model1   
reduced\_model = glm(formula = Y~1, data = data1, family = poisson(link = log))  
lr\_test = anova(reduced\_model, full\_model, test = "Chisq")  
print(lr\_test)

## Analysis of Deviance Table  
##   
## Model 1: Y ~ 1  
## Model 2: Y ~ X  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 9 35.141   
## 2 8 2.482 1 32.659 1.098e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Since p value extremely small, less than alpha = 0.05, reject null hypothesis. Meaning explanatory variable X has significant effect on response variable Y.

1c)

saturated\_MLE = prod(dpois(data1$Y, lambda = data1$Y))  
#note: mean for saturated model is yi/ni, but ni=1 so mean just yi,   
cat("saturated MLE estimate is:" , saturated\_MLE, "\n")

## saturated MLE estimate is: 2.807787e-13

1d)

alpha = coefficients(model1)[1]  
beta = coefficients(model1)[2]  
mean\_xi = exp(alpha + beta \* data1$X)  
full\_model\_MLE = prod(dpois(data1$Y, lambda = mean\_xi))  
cat("full model MLE estimate is:" , full\_model\_MLE, "\n")

## full model MLE estimate is: 8.115431e-14

1e)

mean\_xi2 = exp(alpha)  
reduced\_model\_MLE = prod(dpois(data1$Y, lambda = mean\_xi2))   
print(reduced\_model\_MLE)

## [1] 2.638895e-134

cat("reduced model MLE estimate is:" , reduced\_model\_MLE, "\n")

## reduced model MLE estimate is: 2.638895e-134

1f) (not complete)

null\_di = -2 \* log(saturated\_MLE / reduced\_model\_MLE)  
residual\_di = -2 \* log(full\_model\_MLE / reduced\_model\_MLE)  
cat("null deviance is ", null\_di, "\n")

## null deviance is -557.3497

cat("residual deviance is ", residual\_di, "\n")

## residual deviance is -554.8672

2c)

library(Rsolnp)  
y = c(11, 14, 24, 22, 29)   
ll = function(pi) {  
 ll = log(dmultinom(y, prob = c(pi[1], pi[2], pi[3], pi[4], pi[5])))  
 # return the negative to maximize rather than minimize  
 return(-ll)  
}  
eqn = function(pi) {  
 z1 <- pi[1]-pi[2] # For pi1+pi2+pi3 =1  
 z2 <- pi[3] -pi[4] # For pi1=2pi2 or pi1-2pi2 =0  
 return(c(z1, z2))  
}  
constraints = c(0,0)  
lpi = c(0,0,0,0,0)  
upi = c(1,1,1,1,1)  
pi0 = c(1/5, 1/5, 1/5, 1/5, 1/5)  
ctrl = list(trace = 0)  
sol1 = solnp(pi0, fun = ll, eqfun = eqn, eqB = constraints, LB=lpi,  
 UB=upi, control=ctrl)  
cat("MLe estimate is: ", sol1$pars, "\n")

## MLe estimate is: 0.1238465 0.1238465 0.2278776 0.2278776 0.2873238

2d)

n = sum(y)  
muhat = sol1$pars \* n   
LRT\_2d = sum(((y-muhat)^2)/muhat)  
p\_value\_LRT = 1-pchisq(LRT\_2d, df = length(y) - 1)  
cat("p value is: ", p\_value\_LRT, "\n")

## p value is: 0.9773024

At alpha = 0.05 level of significance, we fail to reject null hypothesis since p value is greater

gsq\_2e = 2 \* sum( y \* log(y/muhat))  
p\_value\_gsq = 1-pchisq(gsq\_2e, df = length(y) - 1)  
cat("p value is: ", p\_value\_gsq, "\n")

## p value is: 0.6803996

At alpha = 0.05 level of significance, we fail to reject null hypothesis since p value is greater

3a)

N = 100  
alpha = 0.05  
pi0 = c(0.15, 0.15, 0.3, 0.4)  
pi = c(0.1, 0.2, 0.3, 0.4)  
crt\_value1 = qchisq(1-alpha, df = length(pi0) - 1)  
lambdaP = N \* sum(((pi-pi0)^2)/pi0)  
power\_P = 1-pchisq(crt\_value1, 3, lambdaP)  
cat("power is ", power\_P)

## power is 0.3026263

3b)

lambdaLRT = 2\*N\*sum(pi\*log(pi/pi0))  
power\_LRT = 1-pchisq(crt\_value1, 3, lambdaLRT)  
cat("power is: ", power\_LRT)

## power is: 0.3080606

3c)

N=0  
power = 0.05   
while(power <0.9){  
 N=N+1  
 lambdaP = N \* sum(((pi-pi0)^2)/pi0)  
 power = 1-pchisq(crt\_value1, 3, lambdaP)  
}  
cat("at N =", N, ", we get a power of ", power)

## at N = 426 , we get a power of 0.9006331

3d)

N2=0  
power2 = 0.05   
while(power2 <0.9){  
 N2=N2+1  
 lambdaLRT = 2\*N2\*sum(pi\*log(pi/pi0))  
 power2 = 1-pchisq(crt\_value1, 3, lambdaLRT)  
}  
cat("at N =", N2, ", we get a power of ", power2)

## at N = 418 , we get a power of 0.9007119

4a)

data2 = data.frame(  
 DoseLevel = c(0, 1, 2, 3, 4, 5), # Convert to factor  
 DeadMoths = c(1, 4, 9, 13, 18, 20)  
)  
#probability of death for x moths dead is x/20  
model2 = glm(cbind(DeadMoths, 20 - DeadMoths) ~ DoseLevel, data = data2, family = binomial)  
summary(model2)

##   
## Call:  
## glm(formula = cbind(DeadMoths, 20 - DeadMoths) ~ DoseLevel, family = binomial,   
## data = data2)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.8186 0.5480 -5.143 2.70e-07 \*\*\*  
## DoseLevel 1.2589 0.2121 5.937 2.91e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 71.138 on 5 degrees of freedom  
## Residual deviance: 1.881 on 4 degrees of freedom  
## AIC: 20.228  
##   
## Number of Fisher Scoring iterations: 4

logistic regression model written on paper using values

4b)

null\_model <- glm(cbind(DeadMoths, 20 - DeadMoths) ~ 1, data = data2, family = binomial)  
summary(null\_model)

##   
## Call:  
## glm(formula = cbind(DeadMoths, 20 - DeadMoths) ~ 1, family = binomial,   
## data = data2)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.1671 0.1832 0.912 0.362  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 71.138 on 5 degrees of freedom  
## Residual deviance: 71.138 on 5 degrees of freedom  
## AIC: 87.485  
##   
## Number of Fisher Scoring iterations: 4

lrt\_test\_4b = anova(null\_model, model2, test = "Chisq")  
lrt\_test\_4b

## Analysis of Deviance Table  
##   
## Model 1: cbind(DeadMoths, 20 - DeadMoths) ~ 1  
## Model 2: cbind(DeadMoths, 20 - DeadMoths) ~ DoseLevel  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 5 71.138   
## 2 4 1.881 1 69.257 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

p value much smaller than even 0.01, so reject null hypothesis, meaning doselevel has significant effect on number of mothsn dead.

4c)

probability\_death = model2$coefficients[1] + model2$coefficients[2]\*3  
cat("probability od death at dose level 3 is: ", probability\_death)

## probability od death at dose level 3 is: 0.9582933

4d) El0.75 is the dose level at which the death level of moths reach 75%, predicted by the full regression model.

lethal\_dose = (log(0.75/(1-0.75)))-model2$coefficients[1]/model2$coefficients[2]  
cat("at a dose level of ", lethal\_dose, "75% of the moths are expected to die")

## at a dose level of 3.337427 75% of the moths are expected to die

4e)

pearson\_residual = resid(model2, type = "pearson")  
hat\_values = hatvalues(model2)  
std\_pearson\_residual = pearson\_residual /sqrt(1-hat\_values)  
cat("standardized pearson residual values are: ", std\_pearson\_residual)

## standardized pearson residual values are: -0.1488074 0.4012877 0.2738542 -0.9064699 -0.03335053 0.8971177

4f)

pearson\_chi\_sq = sum(pearson\_residual^2)  
p\_value = pchisq(pearson\_chi\_sq, df = length(pearson\_residual) -model2$rank, lower.tail = FALSE)  
p\_value

## [1] 0.8598643

since p value much greater than even 0.8 (normally we use a = 0.05), we fail to reject null hypothesis which means that the full model provides a good fit for the data.