Assignment2

Tianyi Mo

September 15,2019

## 1.Introduction

## 1.1 Background

## 1.2 Data

There are four attributtes in the raw data. By looking through the dataset, it is found that every chimpanzee have the same number of instance if data (72 for each chimpanzee), therefore it is not tend to bias towards particular a chimpanzee.

These are the 4 attributes in the raw dataset. actor (1 to 7) condition (0 or 1): prosoc left (0 or 1) pulled left (0 or 1)

## 1.3 Problem:

When human students participate in an experiment like this, they nearly always choose the lever linked to two pieces of food, the prosocial option, but only when another student sits on the opposite side of the table. The question is whether a focal chimpanzee behaves similarly, choosing the prosocial option more often when another animal is present.

load the dataset

dataset = read.delim("assign2.txt", header = TRUE, sep = " ")

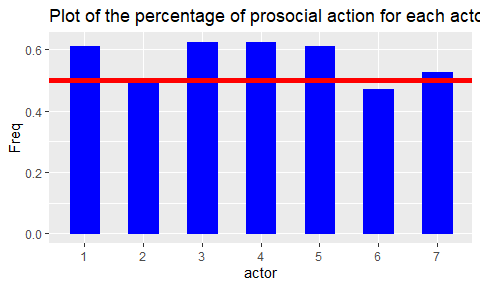
## 2.Create New Attribute

The question is whether a focal chimpanzee behaves similarly, choosing the prosocial option more often when another animal is present. Therefore, I create new attribute “prosocial\_action”, it means whether prosoc\_left and pulled\_left are same. If these two atributes are same, chimpanzee perform prosocially.

dataset['prosocial\_action'] = (dataset$prosoc\_left == dataset$pulled\_left)

## 3.Visualization

library(ggplot2)  
prosocial = xtabs( ~ actor,prosocial\_action, data = dataset)  
prosocial\_percentage = transform(prosocial, Freq=Freq/72)  
  
# generate the bar plot  
ggplot(data=prosocial\_percentage, aes(x=actor, y=Freq))+  
geom\_bar(stat="identity", color="blue", fill="blue",width = 0.5)+  
geom\_hline(yintercept=0.5, linetype=1, color = "red",size = 2)+  
ggtitle("Plot of the percentage of prosocial action for each actor")



## 3.Test dependency of two factors

H0 = prosocial\_action & condition are independent H2 = prosocial\_action & condition are dependent

x = data.frame(xtabs( ~ prosocial\_action + condition, data = dataset))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.2910418

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.5895537

The Chi-square statistics is 0.5895537, it is larger than 0.05, thus it is insignificant and prosocial\_action & condition are independent. However, there might be confounding factors that are ignored and we need to test them.

# 4.Testing confounding factors

actor, prosoc\_left and pulled\_left might be confounding factors

## 4.1 Testing whether actor is confounding factor

H0 : prosocial\_action + condition + actor are independet H1 : prosocial\_action + condition + actor are dependet

#test all 3 factor are independent  
x = xtabs( ~ prosocial\_action + condition + actor, data = dataset)  
model1 = glm(Freq~ prosocial\_action + condition + actor, family = poisson,data = x )  
deviance(model1)

## [1] 9.704807

pchisq(deviance(model1), df = df.residual(model1), lower.tail = FALSE)

## [1] 0.9598556

p-value = 0.9598556, thus prosocial\_action, condition and actor are independent, thus actor is not and confounding a factor.

## 4.2 Testing whether pulled\_left is confounding factor

H0 : prosocial\_action + condition + pulled\_left are independet H1 : prosocial\_action + condition + pulled\_left are dependet

#test all 3 factor are independent  
x = xtabs( ~ prosocial\_action + condition + pulled\_left, data = dataset)  
model1 = glm(Freq~ prosocial\_action + condition + pulled\_left, family = poisson,data = x )  
deviance(model1)

## [1] 1.685609

pchisq(deviance(model1), df = df.residual(model1), lower.tail = FALSE)

## [1] 0.7933302

p-value = 0.7933302, thus prosocial\_action, condition and pulled\_left are independent, thus pulled\_left is not and confounding a factor.

## 4.3 Testing whether prosoc\_left is confounding factor

H0 : prosocial\_action + condition + prosoc\_left are independet H1 : prosocial\_action + condition + prosoc\_left are dependet

#test all 3 factor are independent  
x = xtabs( ~ prosocial\_action + condition + prosoc\_left, data = dataset)  
model1 = glm(Freq~ prosocial\_action + condition + prosoc\_left, family = poisson,data = x)  
deviance(model1)

## [1] 14.43789

pchisq(deviance(model1), df = df.residual(model1), lower.tail = FALSE)

## [1] 0.006020994

p-value = 0.006020994,reject H0, thus 3 factors prosocial\_action, condition and prosoc\_left are dependent.

H0 : prosoc\_left is independent of prosocial\_action and condition H1 : prosoc\_left is dependent of prosocial\_action and condition

#test if prosoc\_left is independent of prosocial\_action and condition   
model2 = glm(Freq~ prosocial\_action \* condition + prosoc\_left, family = poisson,data = x )  
deviance(model2)

## [1] 14.14685

pchisq(deviance(model2), df = df.residual(model2), lower.tail = FALSE)

## [1] 0.002711932

p-value = 0.002711932, reject H0,thus prosoc\_left is dependent of prosocial\_action and condition

H0:prosocial\_action and condition are conditionally independent of prosoc\_left H1:prosocial\_action and condition are conditionally dependent of prosoc\_left

#test if prosocial\_action and condition are conditionally independent of prosoc\_left  
model3 = glm(Freq~ prosocial\_action \* prosoc\_left + condition \* prosoc\_left, family = poisson,data = x )  
deviance(model3)

## [1] 1.443903

pchisq(deviance(model3), df = df.residual(model3), lower.tail = FALSE)

## [1] 0.4858032

p-value = 0.4858032, p-value > 0.05,fail to reject H0,thus prosocial\_action and condition are independent conditional on prosoc\_left

## 5.Conclusion

Question of interest is the relationship between condition and pulled\_left and prosoc\_left, we can combine the data over all 7 chimpanzees.

cta = xtabs( ~ prosoc\_left + condition + pulled\_left, data = dataset)  
##pearson's chisquared test  
summary(cta)

## Call: xtabs(formula = ~prosoc\_left + condition + pulled\_left, data = dataset)  
## Number of cases in table: 504   
## Number of factors: 3   
## Test for independence of all factors:  
## Chisq = 10.877, df = 4, p-value = 0.02798

The chisquare statistic is 10.877, with 4 degrees of freedom and p-value 0.02798. It is less than 0.05 thus is significant. Thus the three factors dependent.

Then test it for each chimpanzee.

## Chimpanzee 1

# generate data for champanzee 1  
data1 = dataset[dataset$actor == 1,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data1))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.2339193

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.6286328

## Chimpanzee 2

Chimpanzee 2 performed a little wired compare to other chimpanzees, it always pull on the left lever and might whatever other factors are. It might be an outlier

# generate data for champanzee 1  
data2 = dataset[dataset$actor == 2,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data2))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] -1.332268e-15

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 1

## Chimpanzee 3

# generate data for champanzee 3  
data3 = dataset[dataset$actor == 3,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data3))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.534171

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.4648585

## Chimpanzee 4

# generate data for champanzee 4  
data4 = dataset[dataset$actor == 4,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data4))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.534171

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.4648585

## Chimpanzee 5

# generate data for champanzee 1  
data5 = dataset[dataset$actor == 5,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data5))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 1.332268e-15

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 1

## Chimpanzee 6

data6 = dataset[dataset$actor == 6,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data6))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.8935135

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.344527

## Chimpanzee 7

data7 = dataset[dataset$actor == 7,]  
x = data.frame(xtabs( ~ prosocial\_action + condition, data = data6))  
  
#using poisson regression to build model  
model = glm(Freq~prosocial\_action + condition, family = poisson,data = x )  
deviance(model)

## [1] 0.8935135

df.residual(model)

## [1] 1

pchisq(deviance(model), df = df.residual(model), lower.tail = FALSE)

## [1] 0.344527

```

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

ctrl + alt + I