Statistical Analysis on the Prosocial Experiment of Chimpanzees

Patrick Tjahjadi

1. Introduction

This report aims to conduct a statistical analysis regarding an experiment evaluating prosocial tendencies of chimpanzees. The experiment involves a chimpanzee sitting at one end of a long table with two levers, one on both left and right sides, called the "focal animal". These levers deliver plates to both ends of the table. However, while the chimpanzee that pulls the lever will always receive a plate that contains food in it, only one side of the two levers will deliver a plate that contains food to the other end of the table. Consider that another chimpanzee is seated at the opposite end of the table, as shown in Figure 1, this situation is called the partner condition.
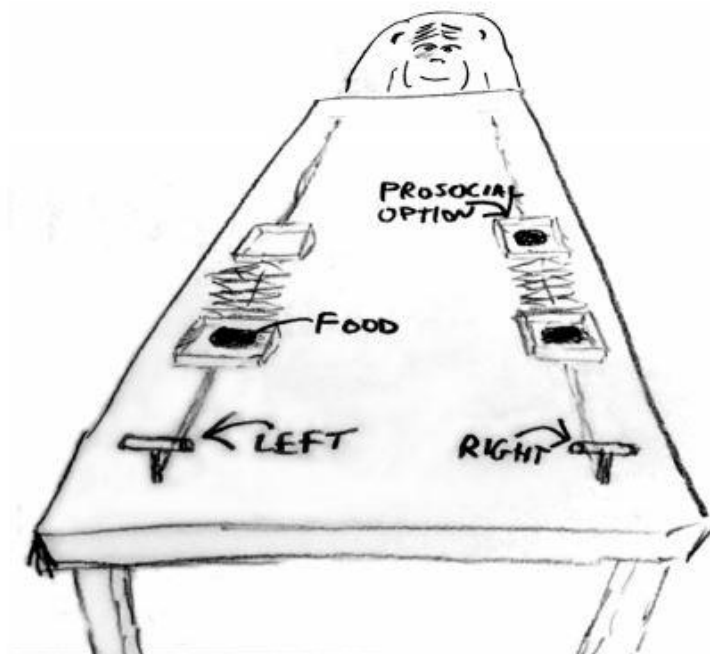

Figure 1: Apparatus for the chimpanzee prosocial experiment from the perspective of the focal animal. Pulling either lever would deliver plates to both ends of the table, but only one side will have food for both sides. The partner condition means that another chimpanzee sits at the other end of the table. The prosocial option would be to pull the lever that contains food for both sides.

There are two conditions for this experiment. The partner condition indicates that another chimpanzee sits at the other end of the table. Otherwise, there is nobody sitting at the other end of the table. The question is whether chimpanzees exhibit prosocial behaviour, that is, whether they choose the prosocial option in the partner condition. This problem will be answered by statistical analysis, more specifically, by providing data visualisation and a statistical model that concludes the prosocial extent of chimpanzees. In this experiment, 7 chimpanzees will be tested.

2. Dataset

The dataset is provided by the University of Melbourne. It has 504 rows representing each experiment. The attributes within the dataset include:

1. actor: The chimpanzee ID, ranging from 1 to 7.
2. condition: Indicator whether the experiment is within the partner condition (1) or not (0).
3. prosoc_left: Indicator of the side with two pieces of food. If the side with two pieces of food is the left side, it is valued 1. Otherwise, it is valued 0.
4. pulled_left: Indicator whether the focal animal pulled the left-hand lever (1) or not (0).

3. Analysis

3.1 Preliminary Model Analysis

First and foremost, it is important to understand the data as a whole first by modelling them.

I suspect that the data is not normally distributed and has signs of heteroskedasticity. Therefore, I will implement a generalised linear model.

Suppose that we observe $Y_i \sim \text{bin}(1, p_i)$, $i = 1, 2, …, 504$. Here, Y indicates pulled_left. Since pulled_left only has 2 classes (either 1 or 0), I will fit a binomial regression model, using a logit link function.

First, the model with only the intercept provides the following summary:

```
> null_model = glm(data$pulled_left ~ 1, family = binomial)
> summary(null_model)

Call:
glm(formula = data$pulled_left ~ 1, family = binomial)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.316  -1.316   1.045   1.045   1.045

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.32017    0.09023   3.548 0.000388 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 685.94  on 503  degrees of freedom
Residual deviance: 685.94  on 503  degrees of freedom
AIC: 687.94

Number of Fisher Scoring iterations: 4
```

Figure 2: Summary of the null model, with only the intercept.

The null model will be considered as baseline, meaning that this is the minimum accuracy my models will be expected to perform, that is, with a deviance of 685.94 and AIC of 687.94. It is expected that my models with predictor variable would have less deviance and AIC.

Including all predictor variables to model pulled_left will yield the following summary:

```
> model = glm(data$pulled_left ~ data$actor + data$condition + data$prosoc_left, family = binomial)
> summary(model)

Call:
glm(formula = data$pulled_left ~ data$actor + data$condition +
    data$prosoc_left, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6482  -1.2414   0.8404   1.0557   1.3505

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.31170    0.24103  -1.293  0.19594
data$actor        0.11525    0.04614   2.498  0.01249 *
data$condition   -0.20183    0.18359  -1.099  0.27161
data$prosoc_left  0.56598    0.18382   3.079  0.00208 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 685.94  on 503  degrees of freedom
Residual deviance: 668.99  on 500  degrees of freedom
AIC: 676.99

Number of Fisher Scoring iterations: 4
```

Figure 3: Summary of the full model.

With a deviance of 668.99 and AIC of 676.99, this model performs better than the null model. However, to further reduce AIC and deviance, a parsimonious model should be formed. Here, I performed stepwise selection to only select the most significant variables:

```
> model2 = step(model, scope = ~.)
Start:  AIC=676.99
data$pulled_left ~ data$actor + data$condition + data$prosoc_left

                   Df Deviance    AIC
- data$condition    1   670.20 676.20
<none>                  668.99 676.99
- data$actor        1   675.30 681.30
- data$prosoc_left  1   678.57 684.57

Step:  AIC=676.2
data$pulled_left ~ data$actor + data$prosoc_left

                   Df Deviance    AIC
<none>                  670.20 676.20
+ data$condition    1   668.99 676.99
- data$actor        1   676.50 680.50
- data$prosoc_left  1   679.76 683.76
```

Figure 4: Stepwise selection of the full model.

Figure 4 suggests that the condition variable is removed. This means that regardless whether it is a partner condition or not, it does not affect the focal animal's tendency to pull the left-hand lever.

Ultimately, the parsimonious model is summarised by the following:

```
> summary(model2)

Call:
glm(formula = data$pulled_left ~ data$actor + data$prosoc_left,
    family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6016  -1.2475   0.8462   1.0197   1.3057

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.41163    0.22333  -1.843   0.0653 .
data$actor         0.11497    0.04608   2.495   0.0126 *
data$prosoc_left   0.56460    0.18359   3.075   0.0021 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 685.94  on 503  degrees of freedom
Residual deviance: 670.20  on 501  degrees of freedom
AIC: 676.2

Number of Fisher Scoring iterations: 4
```

Figure 5: Summary of the model after stepwise selection.

The parsimonious model suggests that actor and prosoc_left are significant predictor variables.

In order to check the fit of the model and deviations from assumptions, I will plot various graphs using R, with the code plot(model2):
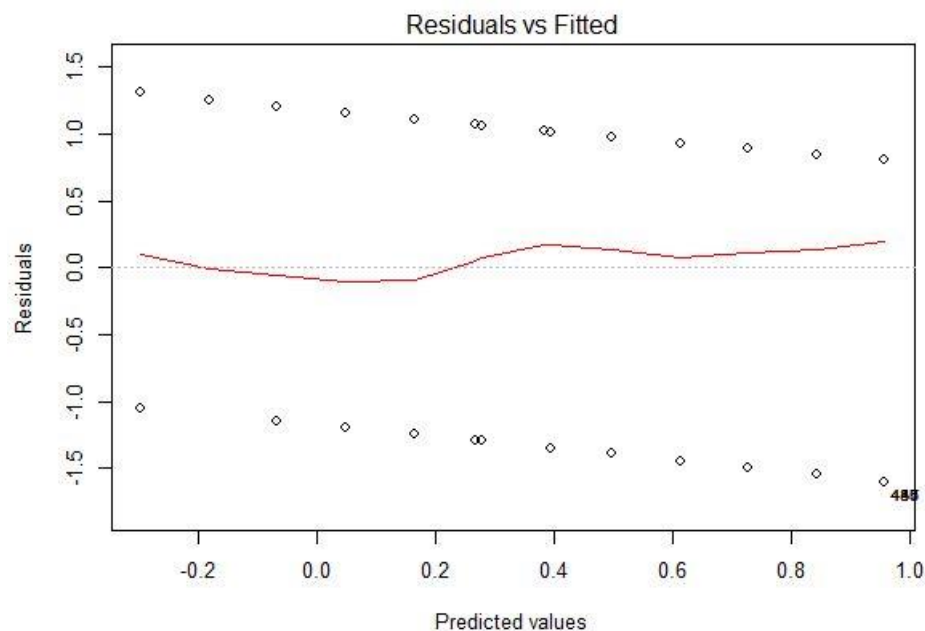

Figure 6: Residuals vs. fitted values plot

Figure 6 shows the trend in the residuals, which indicates bias. Apparently, there is no huge change in pattern of the trend, but there are points with large residuals,

particularly the bottom-right point that has -1.5 residual. Higher residuals indicate more unequal variances.
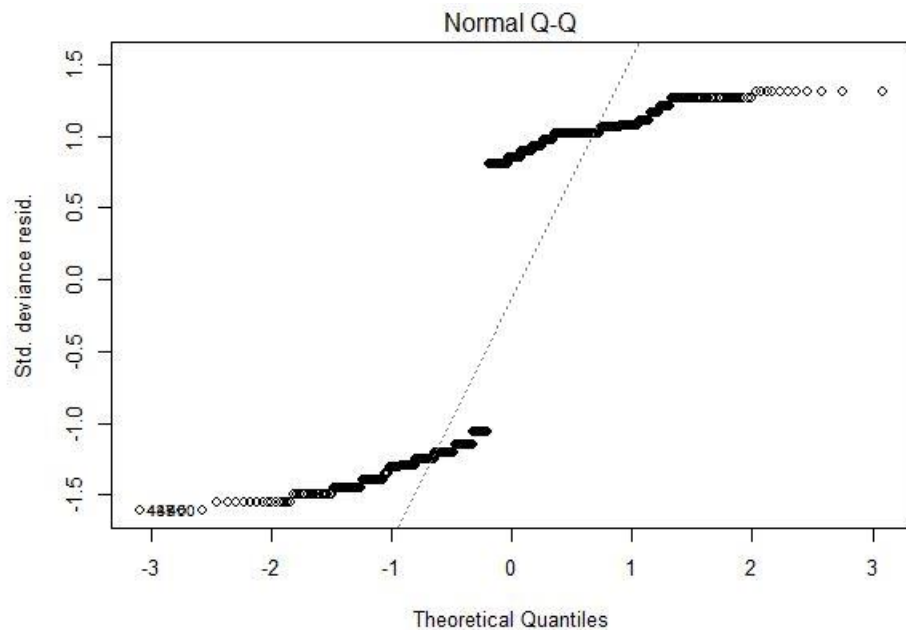

Figure 7: Normal quantile-quantile plot of the standardised residuals.

Figure 7 indicates that there are many points that do not follow the line. This shows that the data is not normally distributed, as expected. Hence, this is the reason why this analysis uses a generalised linear model instead of a simple linear model.
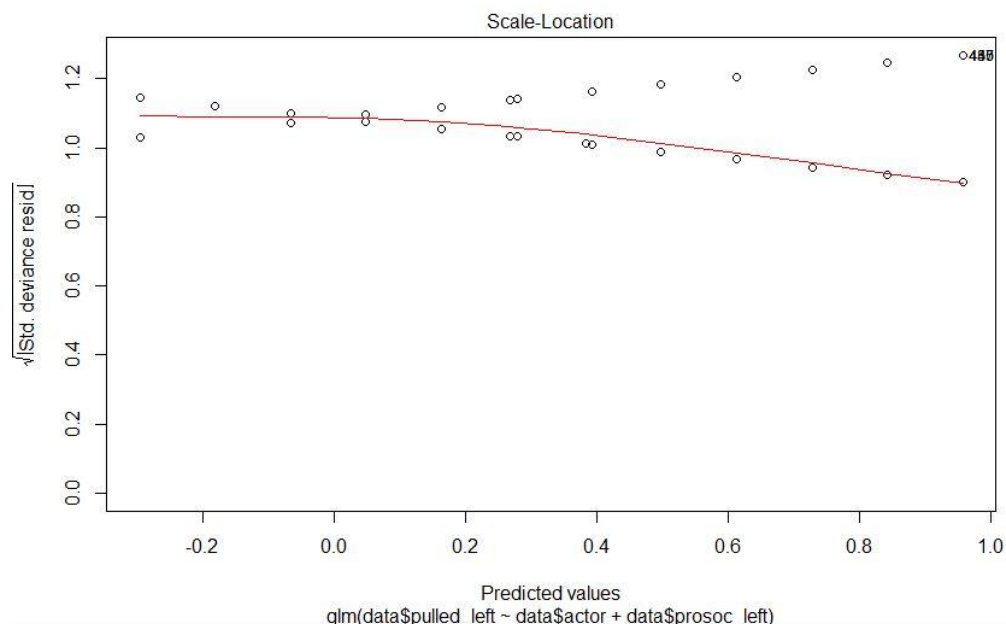

Figure 8: Square roots of absolute values of standardised residuals against fitted values plot.

Figure 8 shows the points with high residual. There is a point of concern, as mentioned in Figure 6, that has a square root of standardised residual of 1.2. The data also shows heteroskedasticity as it gets spread out.
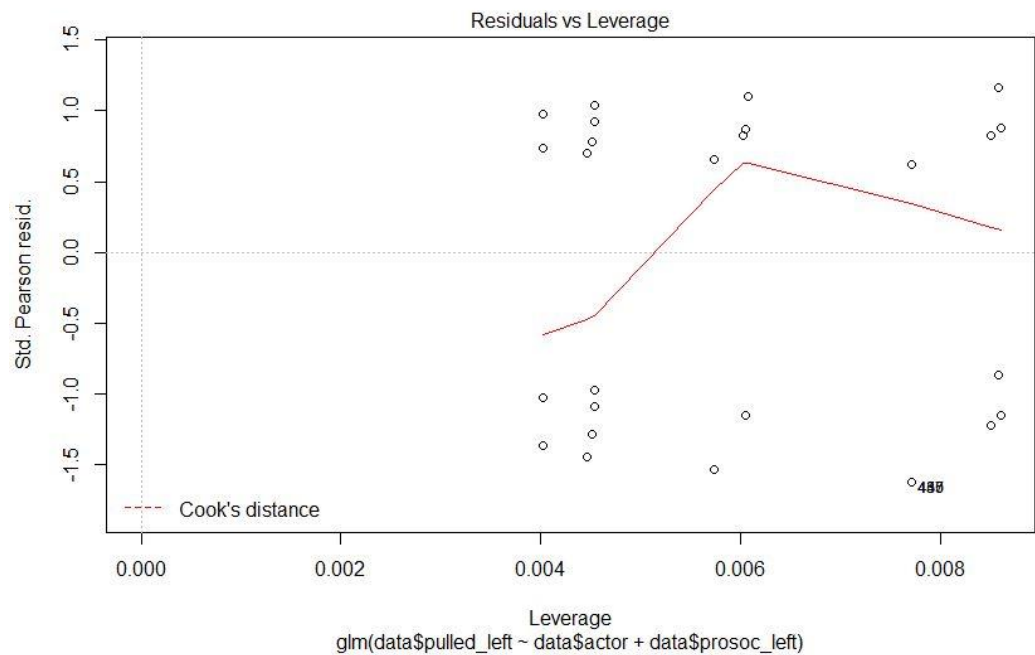

Figure 9: Leverage vs. standardised residuals plot.

Figure 9 shows points that are a poor fit to the model. Since there are is no border for high Cook's distance, it implies that there are no points that have a high Cook's distance. As such, there are no points that are a poor fit to the model.

3.2 Prosocial Dataset

Now, since the dataset has experiments for both the partner condition and non-partner condition, I believe it is best to separate the dataset into two. Since we are only assessing how prosocial these chimpanzees are, I will focus on the entries that fulfils the partner condition.

```
# Separate the dataset by partner condition.
actor2 = c()
prosoc_left2 = c()
pulled_left2 = c()
condition2 = c()
actor3 = c()
prosoc_left3 = c()
pulled_left3 = c()
condition3 = c()
for (i in seq(1,504))
{
  if(data$condition[i] == 0) {
    actor2 = c(actor2, data$actor[i])
    prosoc_left2 = c(prosoc_left2, data$prosoc_left[i])
    pulled_left2 = c(pulled_left2, data$pulled_left[i])
    condition2 = c(condition2, data$condition[i])
  }
  else {
    actor3 = c(actor3, data$actor[i])
    prosoc_left3 = c(prosoc_left3, data$prosoc_left[i])
    pulled_left3 = c(pulled_left3, data$pulled_left[i])
    condition3 = c(condition3, data$condition[i])
  }
}

data_control = data.frame(actor2, condition2, prosoc_left2, pulled_left2)
data_partner = data.frame(actor3, condition3, prosoc_left3, pulled_left3)
```

Figure 10: Code to separate datasets based on the experimental condition.

The dataset happens to be separated evenly: 252 entries for both conditions. Variables ending with 2 are reserved for experiments of the non-partner condition and variables ending with 3 are for experiments of the partner condition.

Next in order to model whether these chimpanzees are prosocial, another column is created. The chimpanzee is considered prosocial if they pulled the lever that has foods for both sides of the table.

```
#Is the chimpanzee being prosocial?
prosocial = c()
for (i in seq(1,252))
{
  if(data_partner$prosoc_left[i] == data_partner$pulled_left[i])
  {
    prosocial = c(prosocial, 1)
  }
  else
  {
    prosocial = c(prosocial, 0)
  }
}
data_partner = data.frame(actor3, condition3, prosoc_left3, pulled_left3, prosocial)
```

Figure 11: Adding another column to assess whether they are prosocial.

3.3 Contingencies

Before fitting a model, it is important to provide initial speculation whether the chimpanzees are being prosocial, briefly. A contingency table sums up the percentage of chimpanzees that are being prosocial within the given data. In order to determine whether they are prosocial, the contingency table provides comparison between the frequencies of the chimpanzees within the variables prosoc_left3, where the left side of the table contains both foods, and pulled_left3, where the chimpanzee decides to pull the left lever.

```
> con_table = table(data_partner$pulled_left3, data_partner$prosoc_left3)
> con_table

    0  1
  0 66 46
  1 60 80
> frac_con_table = prop.table(con_table)
> frac_con_table

            0         1
  0 0.2619048 0.1825397
  1 0.2380952 0.3174603
```

Figure 12: Contingency table, where the rows indicate pulled_left3 and the columns indicate prosoc_left3.

When the left side contains foods for both ends of the table, 80 out of 126 times, the chimpanzees would pull the left lever. This accounts for 31.7% of all trials. Similarly, when the right side contains foods for both ends of the table, 66 out of 126 times, the chimpanzees would pull the right lever, which accounts for 26.1% of all trials. All in all, the chimpanzees are being prosocial at 57.8% of all trials. At first glance, there is not enough evidence to conclude that the chimpanzees are exhibiting prosocial behaviour.

However, we may test if there is a relationship between whether the chimpanzees would pull the lever of the table which side contains both foods (i.e. testing independence between prosoc_left3 and pulled_left3). In order to test for independence, a chi-square test is performed:

```
> # Check for independence
> chisq.test(data_partner$prosoc_left3, data_partner$pulled_left3)

        Pearson's Chi-squared test with Yates' continuity correction

data:  data_partner$prosoc_left3 and data_partner$pulled_left3
X-squared = 5.8018, df = 1, p-value = 0.01601
```

Figure 13: Chi-squared test between prosoc_left3 and pulled_left3

The small p-value (<0.05) indicates that the two variables are not independent. A simple Pearson correlation test may confirm this relationship:

```
> cor.test(data_partner$pulled_left3, data_partner$prosoc_left3,  method = "pearson")

        Pearson's product-moment correlation

data:  data_partner$pulled_left3 and data_partner$prosoc_left3
t = 2.5582, df = 250, p-value = 0.01111
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.03687407 0.27780889
sample estimates:
      cor
0.1597191
```

Figure 14: Pearson Correlation test between prosoc_left3 and pulled_left3

The test shows a slightly positive relationship between prosoc_left3 and pulled_left3. There is a small tendency that the chimpanzees pulled the lever that contains foods for both ends of the table.

3.4 Model Fitting

Similar to the preliminary analysis, I can proceed to fit a binomial regression model, using a logit link function since we are dealing with two levels for the prosocial variable (either they are being prosocial or not at all). This time, in order to expand the analysis for relationships among the variables in the model, interaction variables are introduced. Interaction suggests that the effect of one predictor variable on the response variable also depends on another predictor variable.

```
> #Prosocial model
> prosocial_model = glm(data_partner$prosocial ~ (data_partner$prosoc_left3 + data_partner$actor3
+                        + data_partner$condition3 + data_partner$pulled_left3)^2, family = binomial)
Warning message:
glm.fit: algorithm did not converge
> summary(prosocial_model)

Call:
glm(formula = data_partner$prosocial ~ (data_partner$prosoc_left3 +
    data_partner$actor3 + data_partner$condition3 + data_partner$pulled_left3)^2,
    family = binomial)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-2.409e-06  -2.409e-06   2.409e-06   2.409e-06   2.409e-06

Coefficients: (4 not defined because of singularities)
                                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                                            2.657e+01  9.376e+04   0.000    1.000
data_partner$prosoc_left3                             -5.313e+01  1.089e+05   0.000    1.000
data_partner$actor3                                   -1.163e-10  2.260e+04   0.000    1.000
data_partner$condition3                                       NA         NA      NA       NA
data_partner$pulled_left3                             -5.313e+01  1.174e+05   0.000    1.000
data_partner$prosoc_left3:data_partner$actor3          1.389e-10  2.291e+04   0.000    1.000
data_partner$prosoc_left3:data_partner$condition3             NA         NA      NA       NA
data_partner$prosoc_left3:data_partner$pulled_left3    1.063e+02  9.253e+04   0.001    0.999
data_partner$actor3:data_partner$condition3                   NA         NA      NA       NA
data_partner$actor3:data_partner$pulled_left3          1.249e-10  2.499e+04   0.000    1.000
data_partner$condition3:data_partner$pulled_left3             NA         NA      NA       NA

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3.4297e+02  on 251  degrees of freedom
Residual deviance: 1.4620e-09  on 245  degrees of freedom
AIC: 14

Number of Fisher Scoring iterations: 25
```

Figure 15: Full model of the prosocial model with interaction.

Here, the algorithm to fit the model did not converge. Hence, I will proceed to fit a model without interaction instead.

```
> #Prosocial model
> prosocial_model = glm(data_partner$prosocial ~ (data_partner$prosoc_left3 + data_partner$actor3
+                      + data_partner$condition3 + data_partner$pulled_left3), family = binomial)
> summary(prosocial_model)

Call:
glm(formula = data_partner$prosocial ~ (data_partner$prosoc_left3 +
    data_partner$actor3 + data_partner$condition3 + data_partner$pulled_left3),
    family = binomial)

Deviance Residuals:
    Min      1Q   Median      3Q     Max
-1.5019  -1.2160   0.9466   1.0706   1.1512

Coefficients: (1 not defined because of singularities)
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.28882    0.32582   0.886   0.3754
data_partner$prosoc_left3   0.48032    0.26104   1.840   0.0658 .
data_partner$actor3        -0.03241    0.06491  -0.499   0.6176
data_partner$condition3          NA         NA      NA       NA
data_partner$pulled_left3  -0.13366    0.26533  -0.504   0.6144
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 342.97  on 251  degrees of freedom
Residual deviance: 339.18  on 248  degrees of freedom
AIC: 347.18

Number of Fisher Scoring iterations: 4
```

Figure 16: Full model of the prosocial model with a logit link function.

The Fisher Scoring iteration is 4, which indicates that the Fisher Scoring algorithm need 4 iterations to perform the fit and converge. Now, to create a parsimonious model, stepwise selection is performed. The result is as follows:

```
> prosocial_model2 = step(prosocial_model, scope = ~.)
Start:  AIC=347.18
data_partner$prosocial ~ (data_partner$prosoc_left3 + data_partner$actor3 +
    data_partner$condition3 + data_partner$pulled_left3)


Step:  AIC=347.18
data_partner$prosocial ~ data_partner$prosoc_left3 + data_partner$actor3 +
    data_partner$pulled_left3

                            Df Deviance    AIC
- data_partner$actor3        1   339.43 345.43
- data_partner$pulled_left3  1   339.44 345.44
<none>                           339.18 347.18
- data_partner$prosoc_left3  1   342.60 348.60

Step:  AIC=345.43
data_partner$prosocial ~ data_partner$prosoc_left3 + data_partner$pulled_left3

                            Df Deviance    AIC
- data_partner$pulled_left3  1   339.77 343.77
<none>                           339.43 345.43
- data_partner$prosoc_left3  1   342.89 346.89
+ data_partner$actor3        1   339.18 347.18

Step:  AIC=343.77
data_partner$prosocial ~ data_partner$prosoc_left3

                            Df Deviance    AIC
<none>                           339.77 343.77
- data_partner$prosoc_left3  1   342.97 344.97
+ data_partner$pulled_left3  1   339.43 345.43
+ data_partner$actor3        1   339.44 345.44
```

Figure 17: Stepwise selection of the model with a logit link function.

The Akaike Information Criterion (AIC) was reduced so only significant predictor variables contribute to the model. The result after stepwise selection is as follows:

```
> summary(prosocial_model2)

Call:
glm(formula = data_partner$prosocial ~ data_partner$prosoc_left3,
    family = binomial)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.4196  -1.2181   0.9532   1.1372   1.1372

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.09531    0.17838   0.534   0.5931
data_partner$prosoc_left3   0.45808    0.25702   1.782   0.0747 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 342.97  on 251  degrees of freedom
Residual deviance: 339.77  on 250  degrees of freedom
AIC: 343.77

Number of Fisher Scoring iterations: 4
```

Figure 18: Prosocial model after stepwise selection with a logit link function.

The reduction of deviance compared to the null model is insignificant. There is also barely any reduction of AIC compared to the full model.

The equation to find the probability p that the chimpanzee is being prosocial becomes the following:

$$p = \frac{1}{1 + e^{-\eta}}$$

Equation 1: Probability of a chimpanzee being prosocial using a logit link

Where:

$$\eta = 0.09531 + 0.45808 * prosoc\_left3$$

Equation 2: Estimating η in the logit link function.

Of course, with a p-value of 0.5931 and 0.0747 for the intercept and prosoc_left3, respectively, these two predictor variables are insignificant to model the probability of a chimpanzee being prosocial. Perhaps fitting a binomial regression model with a probit link function would provide better estimates.

```
> #Prosocial model
> prosocial_model = glm(data_partner$prosocial ~ (data_partner$prosoc_left3 + data_partner$actor3
+                       + data_partner$pulled_left3), family = binomial(link = "probit"))
> summary(prosocial_model)

Call:
glm(formula = data_partner$prosocial ~ (data_partner$prosoc_left3 +
    data_partner$actor3 + data_partner$pulled_left3), family = binomial(link = "probit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4900  -1.2248   0.9439   1.0770   1.1585

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.17129    0.20276   0.845   0.3982
data_partner$prosoc_left3     0.29040    0.16193   1.793   0.0729 .
data_partner$actor3          -0.02047    0.04033  -0.508   0.6118
data_partner$pulled_left3    -0.06087    0.16459  -0.370   0.7115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 342.97  on 251  degrees of freedom
Residual deviance: 339.31  on 248  degrees of freedom
AIC: 347.31

Number of Fisher Scoring iterations: 5
```

Figure 19: Full model of the prosocial model with a probit link function.

Again, to create a parsimonious model, stepwise selection is performed.

```
> prosocial_model2 = step(prosocial_model, scope = ~.)
Start:  AIC=347.31
data_partner$prosocial ~ (data_partner$prosoc_left3 + data_partner$actor3 +
    data_partner$pulled_left3)

                            Df Deviance    AIC
- data_partner$pulled_left3  1   339.45 345.45
- data_partner$actor3        1   339.57 345.57
<none>                           339.31 347.31
- data_partner$prosoc_left3  1   342.60 348.60

Step:  AIC=345.45
data_partner$prosocial ~ data_partner$prosoc_left3 + data_partner$actor3

                            Df Deviance    AIC
- data_partner$actor3        1   339.77 343.77
<none>                           339.45 345.45
- data_partner$prosoc_left3  1   342.64 346.64
+ data_partner$pulled_left3  1   339.31 347.31

Step:  AIC=343.77
data_partner$prosocial ~ data_partner$prosoc_left3

                            Df Deviance    AIC
<none>                           339.77 343.77
- data_partner$prosoc_left3  1   342.97 344.97
+ data_partner$actor3        1   339.45 345.45
+ data_partner$pulled_left3  1   339.57 345.57
```

Figure 20: Stepwise selection of the model with a logit link function.

The result after stepwise selection is as follows:

```
> summary(prosocial_model2)

Call:
glm(formula = data_partner$prosocial ~ data_partner$prosoc_left3,
    family = binomial(link = "probit"))

Deviance Residuals:
    Min      1Q   Median       3Q      Max
-1.4196  -1.2181   0.9532   1.1372   1.1372

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.05972    0.11173   0.534   0.5930
data_partner$prosoc_left3   0.28520    0.15969   1.786   0.0741 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 342.97  on 251  degrees of freedom
Residual deviance: 339.77  on 250  degrees of freedom
AIC: 343.77

Number of Fisher Scoring iterations: 4
```

Figure 21: Prosocial model after stepwise selection with a probit link function.

Although it has a residual deviance compared to the logit model, the coefficients have lower p-values, which indicates that they are more significant to model the probability of the chimpanzee being prosocial.

The equation to find the probability p that the chimpanzee is being prosocial becomes the following:

$$p = \Phi(\eta)$$

Equation 3: Probability of a chimpanzee being prosocial using a probit link

Where Φ is the cumulative distribution function of the standard normal distribution and:

$$\eta = 0.05972 + 0.28520 * \text{prosoc\_left3}$$

Equation 4: Estimating η in the probit link function.

With the same Akaike Information Criterion and with both of them having different estimates, it's hard to determine which model should be the final model to predict the probability of a chimpanzee being prosocial. However, I consider the probit model to be the final model since the predictor variables are more significant.

3.5 Goodness-of-fit and Dispersion

A chi-square test is performed to assess whether the model follows a certain distribution.

```
> # Test for GOF
> pchisq(deviance(prosocial_model2), 250, lower.tail = FALSE)
[1] 0.0001342056
```

Figure 22: Chi-square goodness-of-fit test

Since the p-value is very low, we reject H_0 and conclude that the model does not follow a certain distribution. This may be due to overdispersion, and we must test if the model shows signs of overdispersion.

It is important to check for overdispersion since overdispersion reduces the model's F statistic, making the model end up with fewer significant variables. Overdispersion may also cause larger confidence intervals for parameter estimates, which is undesirable. We can check for overdispersion by estimating ϕ, the dispersion parameter. For a binomial model, ϕ should be close to 1:

```
> # Test for overdispersion
> phihat = sum(residuals(prosocial_model2, type="pearson")^2 / 250)
> phihat
[1] 1.008
```

Figure 23: Overdispersion test by estimating ϕ.

The model does not show overdispersion since phihat is close to 1.

Perhaps checking the residual trends might explain the model further. This is done by creating an etahat vs deviance residual plot.

```
> # Check residuals
> plot(residuals(prosocial_model2) ~ predict(prosocial_model2, type="link"), xlab=expression(hat(eta)), ylab = "Deviance residuals")
```
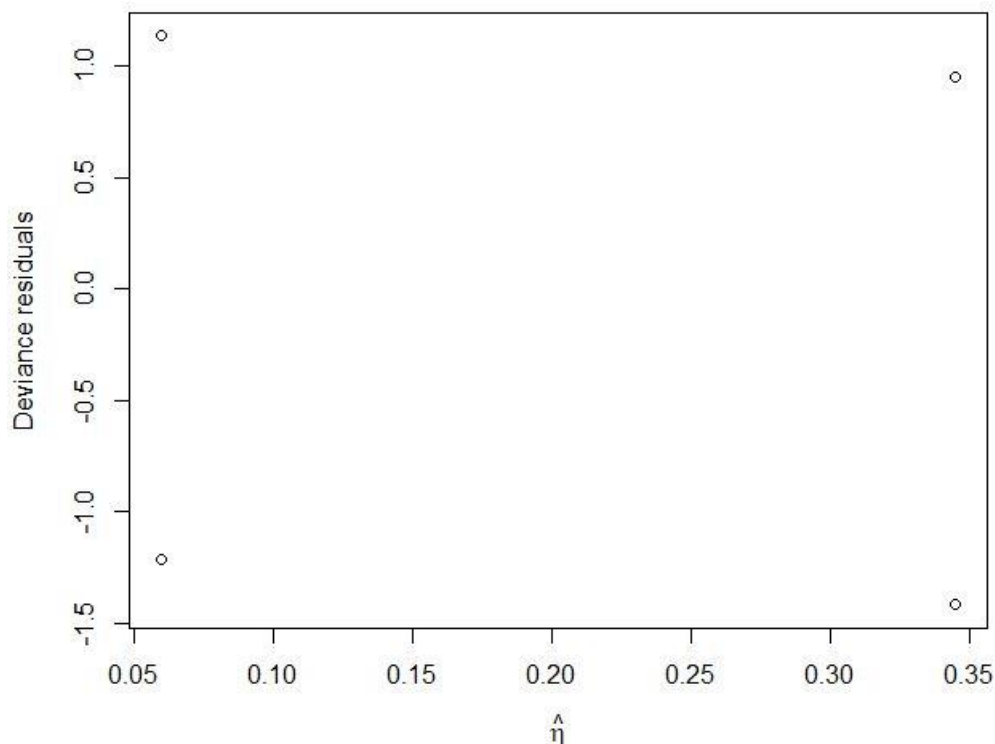

Figure 24: Deviance residuals plot

The plot does follow a trend. As etahat increases, the deviance residuals tend to decrease.

3.6 Model and Data Interpretation

Referring back to Figure 21, it would seem that the only significant predictor variable (besides the intercept) is prosocial_left.

Therefore, in order to determine whether the chimpanzees are prosocial, the only predictor variable to consider is prosocial_left. That is, the probability that a chimpanzee is being prosocial only increases when the left side of the table has both foods. The fact that the pulled_left variable is deemed insignificant indicates that which side the lever they pull does not affect the probability of them being prosocial.

What this means is that it doesn't matter where the side with both foods is. The chimpanzees just tend to pull whichever side they see fit.

By inspection, there also seems to be no pattern whether the chimpanzees tend to pull the lever which side has both foods, regardless of whether it is a partner condition or not. Chimpanzee number 2 also seems to always pull the left lever no matter what, and chimpanzee number 7 almost always pull the left lever. The chimpanzees tend to have random decisions on which lever to pull irrespective of the experiment condition.

4. Conclusion and Future Work

There is no absolute predictor variable that can determine whether the chimpanzees are being prosocial in the experiment or not from the data. Judging by the given data available and the statistical model created, it is insufficient to conclude prosocial behaviour within chimpanzees.

The limitations of the data pose several restrictions on the statistical analysis available. That is, it is infeasible to provide a model with interaction due to the inability of the algorithm to converge. Moreover, there lacks any other predictor variables that may help model the probability of the chimpanzee to being prosocial.

In conclusion, the chimpanzees do not exhibit any prosocial behaviour from the data, and only pulled levers randomly or however they see fit. There are plausible reasons that might affect the tendency for chimpanzees to pull the lever of one side over the other, such as handedness preference of the chimpanzees.