# Task 1: Hypothesis

1. Before forming a hypothesis, import and explore the data.

We want to know whether morning routine affects the amount that an individual exercises. Average daily commute, hours of sleep, and whether an individual has breakfast are the variables affecting morning routine.

1. Adjust columns to make them the appropriate units and classes for each variable used and
2. Initial exploration of the data to inform the hypothesis: plot exercise against the continuous predictors Hours.Slept and avg.commute and check Exercise distribution.

A graph with numbers and dots

Description automatically generatedA graph of numbers and dots

Description automatically generated with medium confidence

Checking the distribution of the response: not normal, but count data, so we can use a Poisson distribution when creating the model.

**Hypothesis**: We expect individuals to exercise more if they sleep longer, eat breakfast, and have a shorter average commute time. If an individual has breakfast,

# Task 2: Model

1. Predictors are hours slept, average commute, and whether an individual eats breakfast. The response is the number of times an individual exercises.
2. Check for correlations or interactions between the predictors:

print(aggregate(data$Hours.Slept, by=list(data$Breakfast), mean, na.rm =T))

## Group.1 x  
## 1 No 7.903333  
## 2 Yes 8.359130# people who eat breakfast may sleep longer.

#Check if continuous variables are correlated.  
cor.test(data$Hours.Slept,data$avg.commute)   
# p = 0.5643, cor = -0.06157292. Not correlated.

1. Create the MAM

The first, most complex model, with all predictors was created and reviewed with a summary and analysis of deviance:

msi1<- glm(Exercise~ Hours.Slept + Hours.Slept\*Breakfast+ I(Hours.Slept^2) + avg.commute, data=data, family= poisson)

anova(msi1, test="Chisq")# p value of average commute = 0.68261 and is the highest p value so we can remove average commute as a predictor. The AIC was 374.32.

Average commute was removed, so now the two final models will confirm whether an interaction term is needed:

m.s<- glm(Exercise~Hours.Slept + Breakfast+ I(Hours.Slept^2), data=data, family= poisson)  
# no interaction term.

m.s.i<- glm(Exercise~Hours.Slept+ Hours.Slept\*Breakfast + I(Hours.Slept^2), data=data, family= poisson)  
# interaction term.

summary() results from two remaining models, and results from analysis of deviance using anova(m.s, m.s.i, test="Chisq") :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | AIC | Residual Deviance | Degrees of Freedom | Deviance | Pr(>Chi) |
| Hours.Slept + Breakfast + I(Hours.Slept^2) | 374.53 | 76.877 | 86 |  |  |
| Hours.Slept + Hours.Slept\*Breakfast + I(Hours.Slept^2) | 372.53 | 72.879 | 85 | 3.9976 | 0.04557 |

1. Checking diagnostic plots + final model

A graph of a line

Description automatically generated with medium confidence A graph of a line

Description automatically generated with medium confidence

The diagnostic plots only look ok for the QQ Normal plot. The variances indicate homoscedasticity, however, the assumptions for glm are less strict than a lm, therefore, we will still use the model with the interaction and quadratic as it is the best fit for the data. It has the lowest AIC, and the interaction term is significant in reducing the deviance of the model. The model and a table of the summary() output are below:

glm(Exercise~Hours.Slept\*Breakfast + I(Hours.Slept^2), data = data, family = poisson)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Standard Error | Z Value | P(>|z|) |
| Intercept | 0.14819 | 1.07038 | 0.138 | 0.8899 |
| Hours.Slept | 0.48511 | 0.27184 | 1.785 | 0.0743 |
| BreakfastYes | -1.31249 | 0.77351 | -1.697 | 0.0897 |
| I(Hours.Slept^2) | -0.04178 | 0.01812 | -2.306 | 0.0211\* |
| Hours.Slept:BrekafastYes | 0.18981 | 0.09693 | 1.958 | 0.0502 |

# Task 3: Results

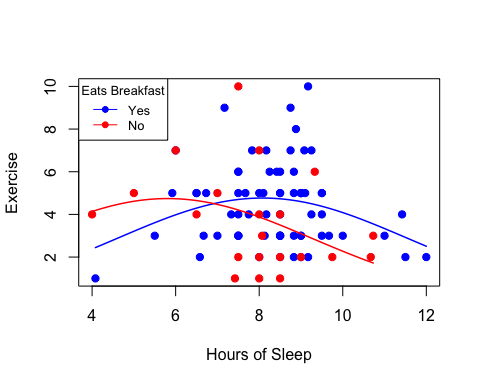
There was a significant quadratic relationship between the the number of times and individual exercised and hours slept (b±SE = -0.04178 ±0.01812; Z-Value 1,85 = -2.3066; p =0.0211). Note that the parameter estimate is on the log scale. While the interaction term of Breakfast was not significant in the model prediction, it did significantly improve the model fit when tested in an analysis of deviance (χ2 1 = 3.9976, p = 0.04557).

# Task 4: Plot

Before plotting, two new data frames were created for the predicted values of the model to ensure a smooth line in the final plot and created two separate lines for each category in Breakfast. The plot is on the original scale after back-transforming from the log transformation which was done when using the poisson family in the glm.

newdata.Y <- data.frame(Breakfast=rep("Yes", 100),  
 Hours.Slept=seq(min(data$Hours.Slept[data$Breakfast=="Yes"]),  
 max(data$Hours.Slept[data$Breakfast=="Yes"]),  
 length.out=100))  
  
newdata.N <- data.frame(Breakfast=rep("No", 100),  
 Hours.Slept=seq(min(data$Hours.Slept[data$Breakfast=="No"]),  
 max(data$Hours.Slept[data$Breakfast=="No"]),  
 length.out=100))  
  
predicted.Y2 <- predict(m.s.i, newdata.Y, type='response')  
predicted.N2 <- predict(m.s.i, newdata.N, type='response')

plot(Exercise ~ Hours.Slept, data=data, pch=NA, xlab="Hours of Sleep", ylab="Exercise")  
points(Exercise ~ Hours.Slept, data=data[data$Breakfast=="Yes", ], pch=19, col="blue")   
points(Exercise ~ Hours.Slept, data=data[data$Breakfast=="No", ], pch=19, col="red")  
lines(predicted.Y2[order(newdata.Y$Hours.Slept)] ~  
 sort(newdata.Y$Hours.Slept), lwd=1.5, col="blue")  
lines(predicted.N2[order(newdata.N$Hours.Slept)] ~  
 sort(newdata.N$Hours.Slept), lwd=1.5, col="red")  
legend(x="topleft", legend=c("Yes", "No"), pch=19,  
col=c("blue", "red"), lwd=c(1,1), title="Eats Breakfast", cex=0.8)



**Figure 1. Effect of morning routing on exercise**.

Hours of sleep have a quadratic effect on the amount an individual exercise. While there was no significant effect of breakfast on the relationship between sleep and exercise, the model suggested an interaction (Z-Value 1,85 = -2.3066; p = 0.0502). At the recommended 8 hours of sleep for adults who eat breakfast, the predicted number of times exercised is 4.76. Adults who don’t eat breakfast exercise more with less sleep, and at the recommended 8 hours their predicted exercise is 3.88.