Jorō Spiders and Traffic


February 29, 2024

People in the south-east of the US have been getting excited by the presence of *Trichonephila clavata*, the Jorō spider. It is invasive, but it is noteworthy the authors of the paper said these appear mainly not problematic–so far–for native spiders `https://www.mdpi.com/2813-3323/2/1/4`. Part of the reason is that they (the spiders, I don't know about the authors) like cities and noise and chaos more than most spiders that rely on movement in their webs to detect lunch. Their Table 1 summarizes their data, and is appropriate for either a multilevel binary logistic regression or a non-multilevel (non-binary) logistic regression.[1] They found webs on 20 fairly busy streets and used a tuning fork to touch the web to simulate an insect. They did this about 20 times for each street. If they had insect variables they would have used a multilevel model, but in this table they did not (I have only looked at the tables and photos of spiders, not the whole paper). They recorded whether the Jorō attacked the tuning fork or retreated/did nothing. Here are some of their data:

```
#Traffic Density (1 to 4 scale, 4 most dense)
#or cars per day (cpd)
cpd <- c(20,25,30,30,50,70,90,250,270,380,620,2080,
         4040,4910,6060,10200,10200,10500,12700,45100)
tdcat <- c(rep(1,7),rep(2,4),rep(3,4),rep(4,5))
trials <- c(18,18,15,17,17,15,19,20,20,18,17,
            26,18,19,24,7,21,12,16,22)
attacks <- c(17,13,12,5,12,5,11,16,14,9,14,14,
             9,11,7,2,13,6,11,12)
avespidmass <- c(.33,.55,.37,.34,.23,.66,.40,
     .42,.40,.53,.36,.64,.35,.58,.38,.22,.52,
     .45,.50,.28)
```

[1] They did not account for area in their model, but .... They also could have used individual spider variables ....

- Why not just use the proportions? (because that would weight the 7 trial street and the 22 trial street the same)

When conducting a logistic regression like this you need to tell R both the number of trials and the number of one of the categories, here attacks. The thing you predict on the left side of the $\sim$ is composed of two variables: the success and failures (or here, attacks and non-attacks.

```
glm(cbind(attacks,trials-attacks) ~
```

Their Table 3 showed the results of two models. Predicting this from spider mass and the log of the traffic density, and from spider mass and volume category.

btw, here is the effect of logging.

```
library(e1071)
skewness(cpd); skewness(log(cpd))

## [1] 2.826108
## [1] 0.0251091

par(mfrow=c(2,2))
hist(cpd); qqnorm(cpd); qqline(cpd)
hist(log(cpd)); qqnorm(log(cpd)); qqline(log(cpd))
```

Here is their model with the logged traffic density. I also checked if there was an interaction.

```
summary(m1 <- glm(cbind(attacks,trials-attacks) ~
        log(cpd) + avespidmass,family=binomial))

##
## Call:
## glm(formula = cbind(attacks, trials - attacks) ~ log(cpd) + avespidmass,
##     family = binomial)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.35369    0.49479   2.736  0.00622 **
## log(cpd)    -0.10214    0.04466  -2.287  0.02218 *
## avespidmass -0.69242    0.89932  -0.770  0.44134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 50.90  on 19  degrees of freedom
## Residual deviance: 44.81  on 17  degrees of freedom
## AIC: 113.31
##
## Number of Fisher Scoring iterations: 4
```

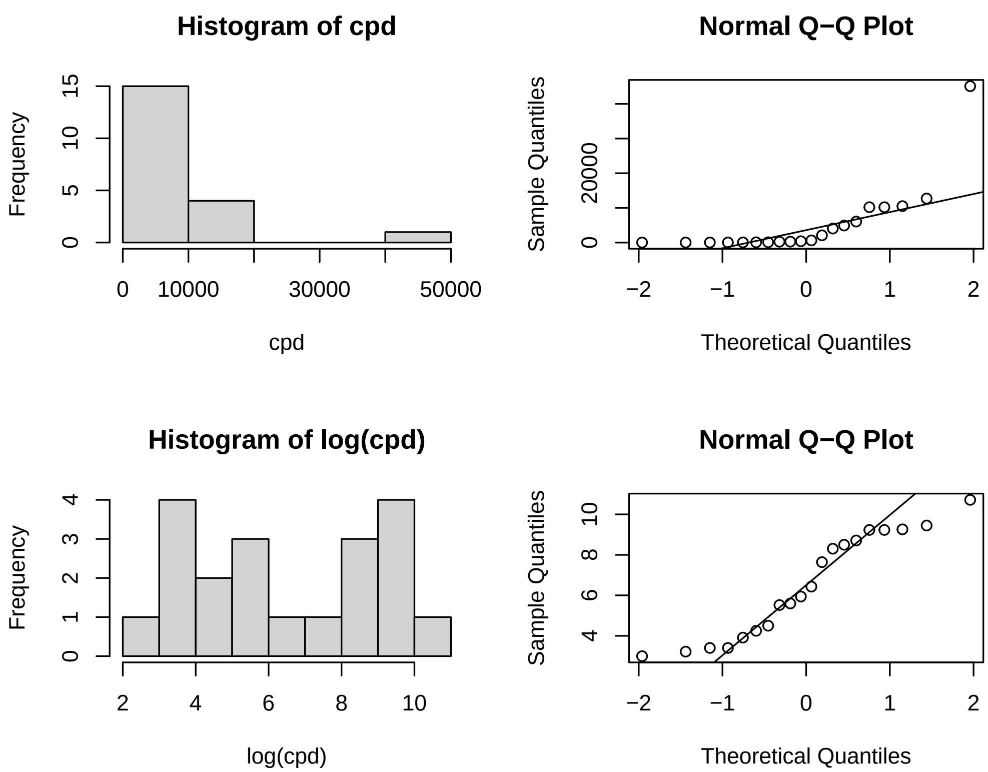

Figure 1: Histograms and QQ plots for the traffic density and the log of it.

```
anova(m1,m2 <- update(m1, .~. + log(cpd):avespidmass),test="Chisq")

## Analysis of Deviance Table
##
## Model 1: cbind(attacks, trials - attacks) ~ log(cpd) + avespidmass
## Model 2: cbind(attacks, trials - attacks) ~ log(cpd) + avespidmass + log(cpd):avespidmass
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        17     44.810
## 2        16     41.221  1   3.5892  0.05816 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction suggests that spider mass might be important, so they maybe should have looked at differences among the individual spiders, but I am not a spider person.

```
preds <- predict(update(m1, .~. - avespidmass),type="response")
plot(log(cpd),preds,type='b',las=1,ylim=c(0,1),
     xlab="log traffic",ylab="Prob. attack")
```

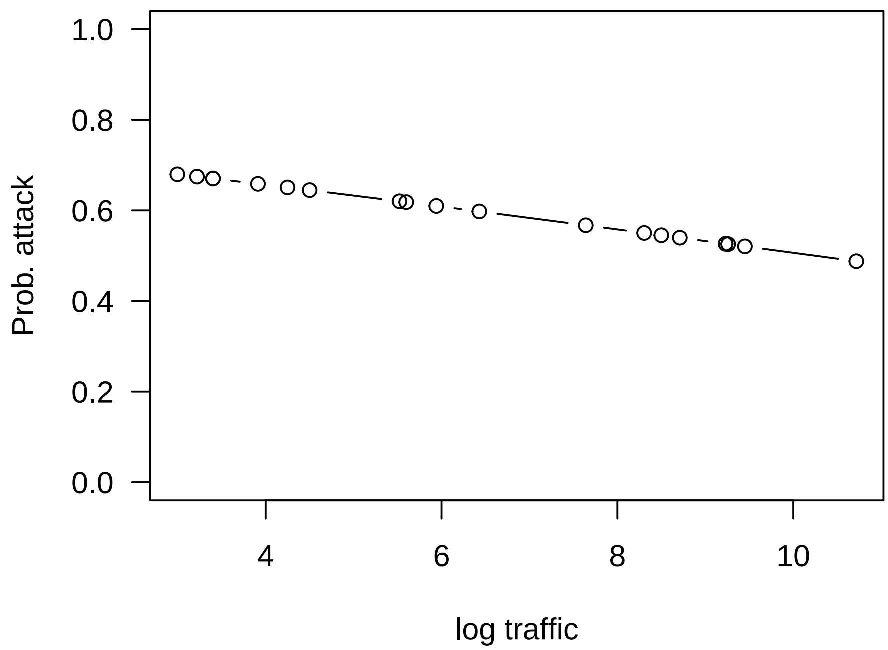

Figure 2: Predicting Jorō spiders attacking the tuning fork.

Bridge to Multilevel

Suppose that we did have trial (or individual Jorō spider data rather than something like average weight, which pretty odd to have used). Let's pretend there is a variable `tuning` that relates to how much the web is shaken and therefore is predicted to positively relate to the probability of attack, unless it is really high (above 4).

```
tuning <- rnorm(sum(trials))
```

Make the data long. These are made up!

```
set.seed(383838)
cpdLong <- rep(cpd,trials)
street <- LETTERS[rep(1:length(trials),trials)]
streetef <- rep(runif(length(cpd),-1,1),trials)
attack <- rbinom(length(street),1,plogis(streetef-scale(log(cpdLong))+tuning))
longdata <- data.frame(street,cpdLong,attack)
```

```r
library(lme4)
```

```r
m0 <- glmer(attack ~ 1 + (1|street),family=binomial)
summary(m0)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: attack ~ 1 + (1 | street)
##
##      AIC      BIC   logLik deviance df.resid
##    441.8    449.6   -218.9    437.8      357
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1992 -0.7524  0.2855  0.7339  2.6027
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  street (Intercept) 1.566    1.251
## Number of obs: 359, groups:  street, 20
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.1468     0.3080   0.477    0.634

m1 <- update(m0, .~. + tuning + log(cpdLong))
anova(m0,m1)

## Data: NULL
## Models:
## m0: attack ~ 1 + (1 | street)
## m1: attack ~ (1 | street) + tuning + log(cpdLong)
##    npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## m0    2 441.79 449.56 -218.9   437.79
## m1    4 399.60 415.13 -195.8   391.60 46.191  2  9.326e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(m1)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: attack ~ (1 | street) + tuning + log(cpdLong)
##
##      AIC      BIC   logLik deviance df.resid
##    399.6    415.1   -195.8    391.6      355
```

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9177 -0.6511  0.2621  0.6077  2.9105
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  street (Intercept) 0.273    0.5225
## Number of obs: 359, groups:  street, 20
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.09117    0.51604   5.990 2.10e-09 ***
## tuning        0.72459    0.14799   4.896 9.76e-07 ***
## log(cpdLong) -0.44947    0.07386  -6.085 1.16e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) tuning
## tuning       0.204
## log(cpdLng) -0.941 -0.187
```

```
confint(m1,method="Wald")
```

```
##                   2.5 %     97.5 %
## .sig01               NA         NA
## (Intercept)   2.0797470  4.1025838
## tuning        0.4345433  1.0146434
## log(cpdLong) -0.5942383 -0.3047088
```

```
anova(m1,update(m1, .~. + tuning:log(cpdLong)))
```

```
## Data: NULL
## Models:
## m1: attack ~ (1 | street) + tuning + log(cpdLong)
## update(m1, . ~ . + tuning:log(cpdLong)): attack ~ (1 | street) + tuning + log(cpdLong) + tunin
##                                         npar    AIC    BIC  logLik deviance
## m1                                         4 399.60 415.13 -195.80   391.60
## update(m1, . ~ . + tuning:log(cpdLong))    5 401.58 420.99 -195.79   391.58
##                                          Chisq Df Pr(>Chisq)
## m1
## update(m1, . ~ . + tuning:log(cpdLong)) 0.0244  1     0.8758
```

```
anova(m1,update(m1, .~. - (1|street) + (tuning|street)))
```

```
## Data: NULL
## Models:
## m1: attack ~ (1 | street) + tuning + log(cpdLong)
```

```
## update(m1, . ~ . - (1 | street) + (tuning | street)): attack ~ tuning + log(cpdLong) + (tuning
##                                                       npar    AIC    BIC logLik
## m1                                                       4 399.60 415.13 -195.8
## update(m1, . ~ . - (1 | street) + (tuning | street))    6 395.39 418.69 -191.7
##                                                       deviance  Chisq Df
## m1                                                      391.60
## update(m1, . ~ . - (1 | street) + (tuning | street))   383.39 8.2083  2
##                                                       Pr(>Chisq)
## m1
## update(m1, . ~ . - (1 | street) + (tuning | street))     0.0165 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3 shows the predicted values by the two predictor variables.

```
par(mfrow=c(1,2))
plot(tuning,predict(m1,type="response",re.form=NA),
     ylab="Prob Attack",las=1)
plot(log(cpdLong),predict(m1,type="response",re.form=NA),
     ylab="Prob Attack",las=1,xlab="ln(traffic)")
```

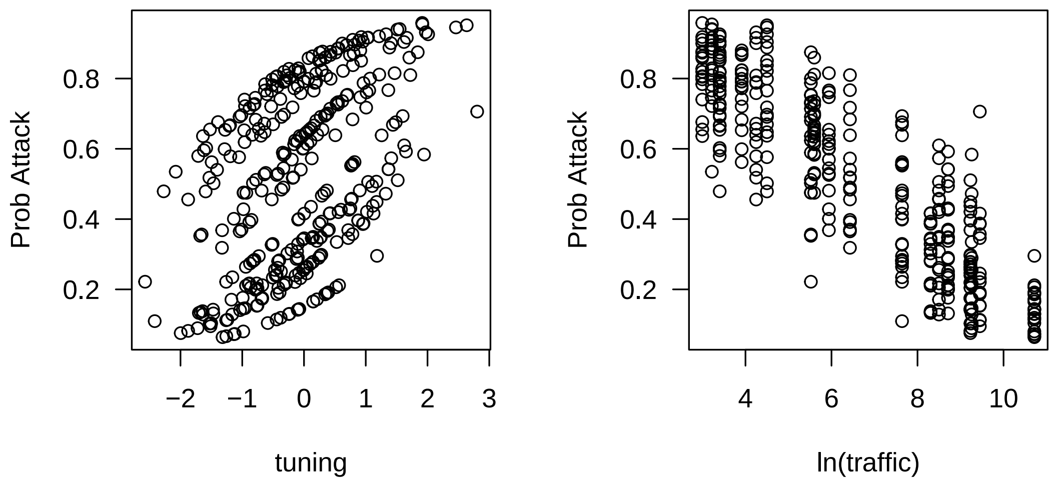

Figure 3: Probability of attack by the tuning variable and traffic for Jorō spiders.