Assignment5Mrkdwn

2023-05-10

Observer should be included as a random effect because different observers may introduce variance into the data due to factors such as the step size of the observer or their interpretation of what is meant by a weed being “exactly” at the tip of their shoe.

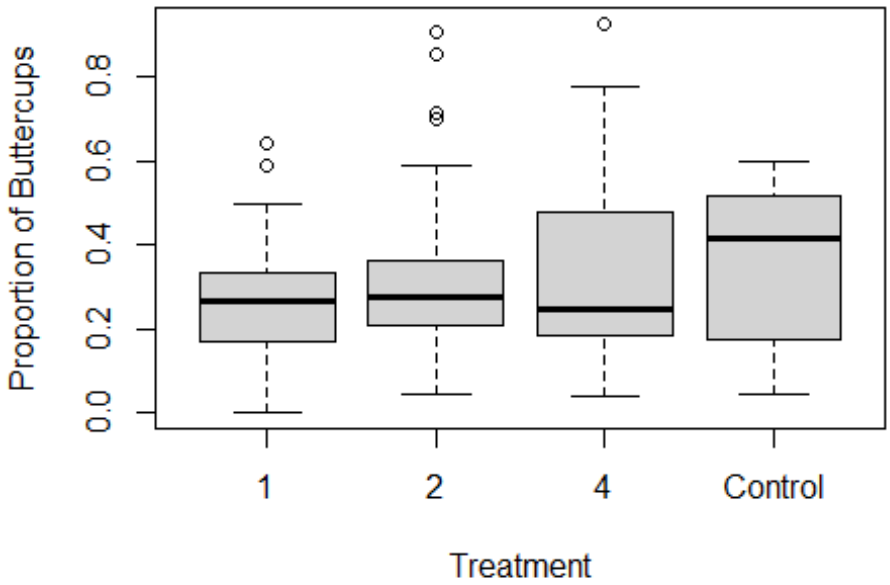
A binomial distribution is appropriate because it allows us to model the proportion of successes (steps with butter cup present) out of the total number of steps. This allows us to find the probobility of finding a buttercup after each step.

```
library(lme4)

## Loading required package: Matrix

data <- read.csv("initobs.csv")
data$Trt <- factor(data$Trt)
data$Farm <- factor(data$Farm)

y <- cbind(data$Bc, data$NotBc)
plot(x=data$Trt, y=data$Bc/data$NotBc, ylab="Proportion of Buttercups",
     xlab="Treatment")
```



```r
mod1 <- glmer(y ~ Trt + Farm + (1|Observer),data = data, family = binomial)
mod2 <- glmer(y ~ Trt + (1|Observer) + (1|Farm),data = data, family =
binomial)
mod3 <- glmer(y ~ Trt + (1|Observer) + (1|Farm:Plot),data = data, family =
binomial)

anova(mod1, mod2, mod3)
```

```
## Data: data
## Models:
## mod2: y ~ Trt + (1 | Observer) + (1 | Farm)
## mod3: y ~ Trt + (1 | Observer) + (1 | Farm:Plot)
## mod1: y ~ Trt + Farm + (1 | Observer)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod2    6 482.23 498.16 -235.12   470.23
## mod3    6 478.51 494.44 -233.26   466.51 3.7174  0
## mod1    9 477.48 501.36 -229.74   459.48 7.0362  3    0.07075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(mod1) #Best model
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: y ~ Trt + Farm + (1 | Observer)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##    477.5    501.4   -229.7    459.5       96
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.37497 -0.88004  0.02078  0.67377  3.02819
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Observer (Intercept) 0.04433  0.2105
## Number of obs: 105, groups:  Observer, 5
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.30669    0.18257  -7.157 8.24e-13 ***
## Trt2         0.15889    0.12930   1.229   0.2191
## Trt4         0.21855    0.12847   1.701   0.0889 .
## TrtControl   0.22587    0.15760   1.433   0.1518
## FarmB        0.19313    0.19957   0.968   0.3332
## FarmF       -0.31136    0.20704  -1.504   0.1326
## FarmR       -0.39467    0.21223  -1.860   0.0629 .
## FarmS        0.05252    0.19425   0.270   0.7869
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Trt2   Trt4   TrtCnt FarmB  FarmF  FarmR
## Trt2        -0.356
## Trt4        -0.355  0.525
## TrtControl -0.300  0.427  0.430
## FarmB       -0.598 -0.004 -0.009  0.006
## FarmF       -0.567 -0.021 -0.022 -0.002  0.723
## FarmR       -0.555 -0.015 -0.020 -0.011  0.707  0.679
## FarmS       -0.605 -0.012 -0.016 -0.006  0.767  0.736  0.720
```

Model1 which includes observer as a random effect and treatment and farm as fixed effects is the best model. This is because it has the lowest AIC meaning it has the best tradeoff between goodness of fit and model complexity.

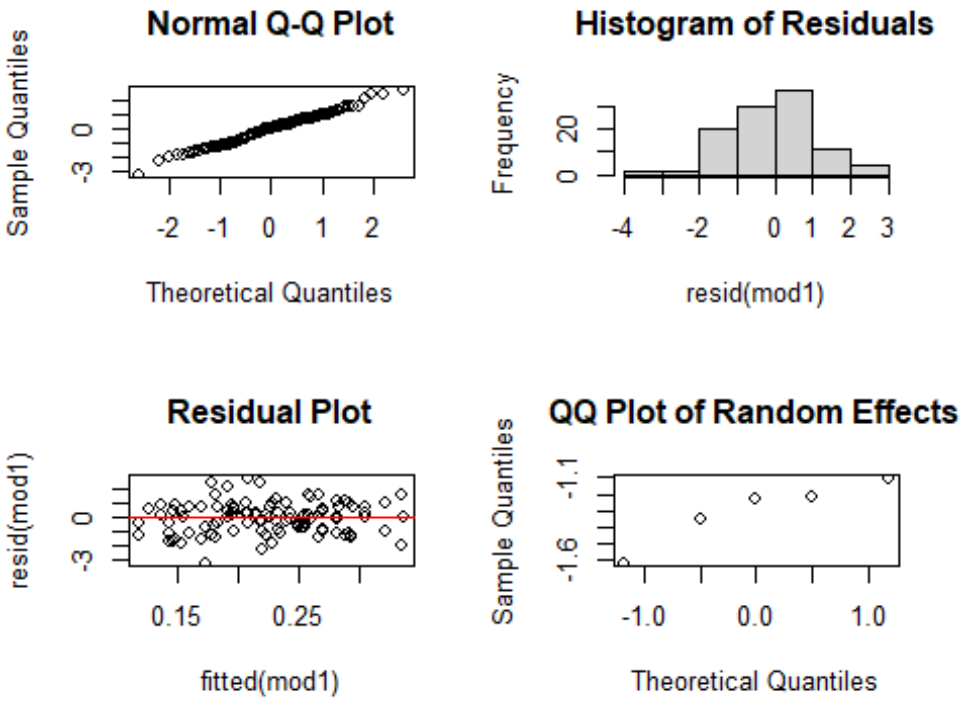
The scaled residuals look good because the median is close to 0 (0.0208) and the 1st and 3rd quartile are reasonably symetric.

The random effect coefficient of 0.044 suggests there are differences between observers when determining if a buttercup is found at the end of a step. This introduces variance into our data which we need to account for.

The fixed effects all have insignificant p-values. This means there is no significant difference between treatments in each farm. The regression equation is log odds of finding buttercup = -1.307 + (0.159 x trt2) + (0.219 x trt4) + (0.226 x control) + (0.193 x farmB) + (-0.311 x farmF) + (-0.395 x farmR) + (0.053 x farmS).

my overall conclusion is that there is no significant difference between treatments in the farms. Our original plot did not show a clear difference between the groups. This observation was backed up by generalised linear mixed effects models coefficients.

```
par(mfrow=c(2,2))
qqnorm(resid(mod1), main="Normal Q-Q Plot")
hist(resid(mod1), main="Histogram of Residuals")
plot(fitted(mod1), resid(mod1), main = "Residual Plot")
abline(0,0, col='red')
qqnorm(coef(mod1)$Observer[,1], main="QQ Plot of Random Effects")
```



The residuals appear to be normally distributed as the residuals can be fit to a straight line in our Normal Q-Q plot and create a bell curve shape in our Histogram of Residuals. The residuals are scattered above and bellow 0 in the Residual plot so we can assume there is constant variance between the residuals.

The points in the QQ Plot of Random Effects can be fit to a striaght line. This menas we can assume our random effects are normally distributed.