Assignment 5 STAT 315-463: Multivariable Statistical Methods and Applications

Lisa Lu 31088272

```
# Read in data file
initobs <- read.table("initobs.csv", header = TRUE, sep = ',', na.strings = "na")
plot <- as.factor(initobs$Plot)
trt <- as.factor(initobs$Trt)
```

**a) Explain why observer should be included in a model as a random effect.**

From the left plot below, we can see that the number of occurrence of each observer varies greatly. The right plot also suggested that the distribution of the percentage of ground covered by the weed differs from each observer. Therefore, the observers are the source of random variation, which should be taken into consideration when choosing the model as it brings potential effects and may influence the relationship between the control of the weed and the different treatments.

```
pardef <- par()
par(mfrow = c(1,2))
plot(as.factor(initobs$Observer), xlab="Observer", ylab="Count")
plot(as.factor(initobs$Observer),initobs$Bc/initobs$Steps * 100,
     xlab="Observer", ylab="% of ground cover of the weed")
```
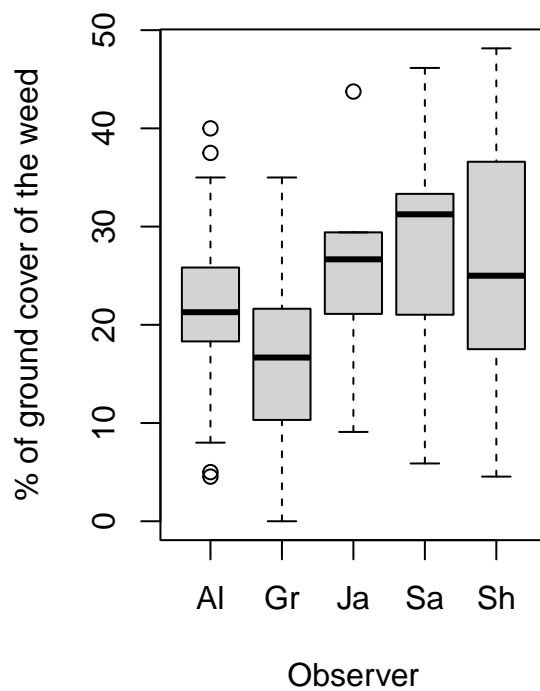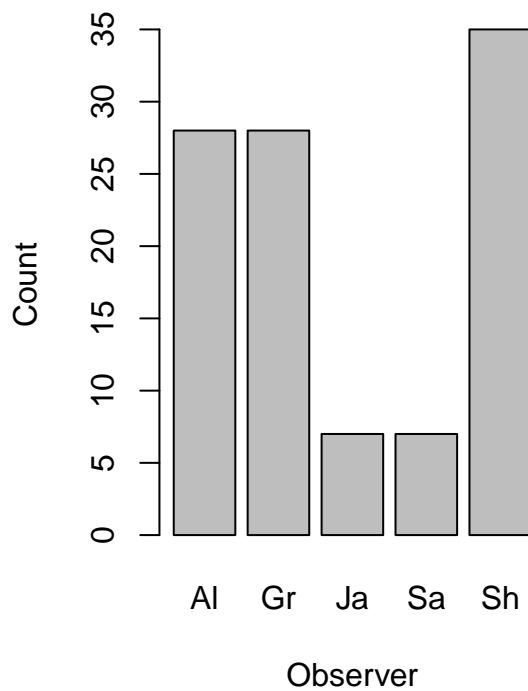
b) What distribution would be appropriate for these data and why? From the histogram of the frequency of the buttercup, we can see that it does not follow the normal distribution. Furthermore, because it is the count data of how frequent the buttercup occurs, therefore, this dataset follows the binomial distribution.

```
hist(initobs$Bc)
```

Histogram of initobs$Bc
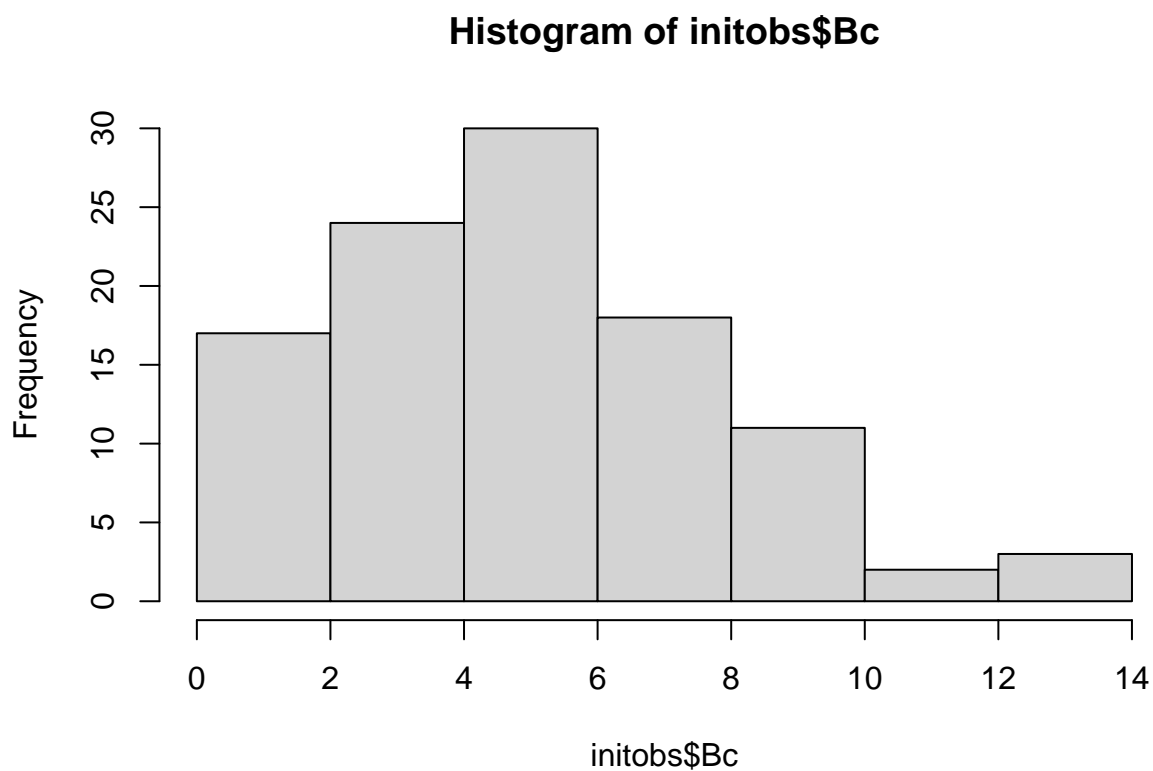

```
pardef <- par()
par(mfrow = c(2,2))

plot(as.factor(initobs$Trt), initobs$Bc/initobs$Steps,
     xlab="Treatment",ylab="% of ground cover of the weed")
plot(as.factor(initobs$Farm), initobs$Bc/initobs$Steps,
     xlab="Farm", ylab="% of ground cover of the weed")
plot(as.factor(initobs$Plot), initobs$Bc/initobs$Steps,
     xlab="Plot", ylab="% of ground cover of the weed")
plot(as.factor(initobs$Observer),initobs$Bc/initobs$Steps,
     xlab="Observer", ylab="% of ground cover of the weed")
```
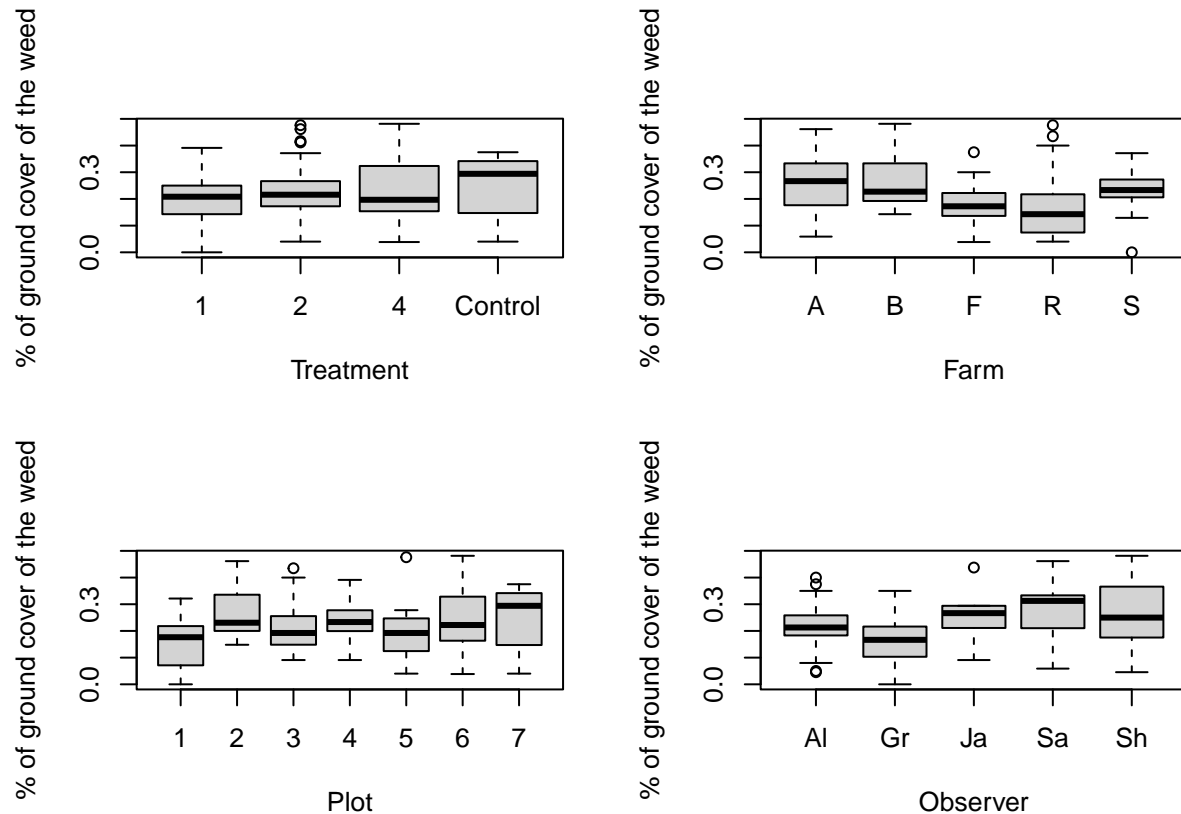
These data are not continuous, therefore it is not very suitable to use normal linear regression models. The underlying relationship between the response variable and the predictor variables is not linear here. The graphs above suggest that Because the buttercup after the treatment will either appear or not appear. Therefore, this dataset is following binomial distribution.

c) Fit an appropriate random effects model to these data

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
Trt <- as.factor(initobs$Trt)
m1 <- glmer(cbind(Bc, NotBc) ~ trt + (1 |Observer) + (1|Farm),
            family = binomial, data = initobs)
```

```
summary(m1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(Bc, NotBc) ~ trt + (1 | Observer) + (1 | Farm)
##     Data: initobs
##
##      AIC      BIC   logLik deviance df.resid
##    482.2    498.2   -235.1    470.2       99
```

4

```
##
## Scaled residuals:
##     Min      1Q   Median      3Q      Max
## -2.3352 -0.7768  0.0200  0.6535  2.7951
##
## Random effects:
##  Groups    Name          Variance Std.Dev.
##  Observer (Intercept) 0.05627  0.2372
##  Farm      (Intercept) 0.04474  0.2115
## Number of obs: 105, groups:  Observer, 5; Farm, 5
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.3821     0.1767  -7.821 5.25e-15 ***
## trt2          0.1564     0.1291   1.211   0.2259
## trt4          0.2165     0.1283   1.687   0.0915 .
## trtControl    0.2235     0.1574   1.420   0.1556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) trt2   trt4
## trt2       -0.378
## trt4       -0.380  0.525
## trtControl -0.313  0.427  0.430
```

- a. Discuss the results of the analysis, include comments about the following:

  - i). Scaled residuals
  - ii). Random effects
  - iii). Fixed effects

- b. What would your overall conclusion be?

d) Draw graphs of the residuals and the random effects and comment on these.

```
# Create plots of the residuals
par(mfrow=c(1,3))
qqnorm(resid(m1, main = "Residual Plot"))
abline(0,1)
hist(resid(m1), main = "Histogram of Residuals")
plot(fitted(m1), resid(m1), xlab = "Fitted", ylab="Residuals")
abline(0,0)
```

Normal Q–Q Plot   **Histogram of Residuals**
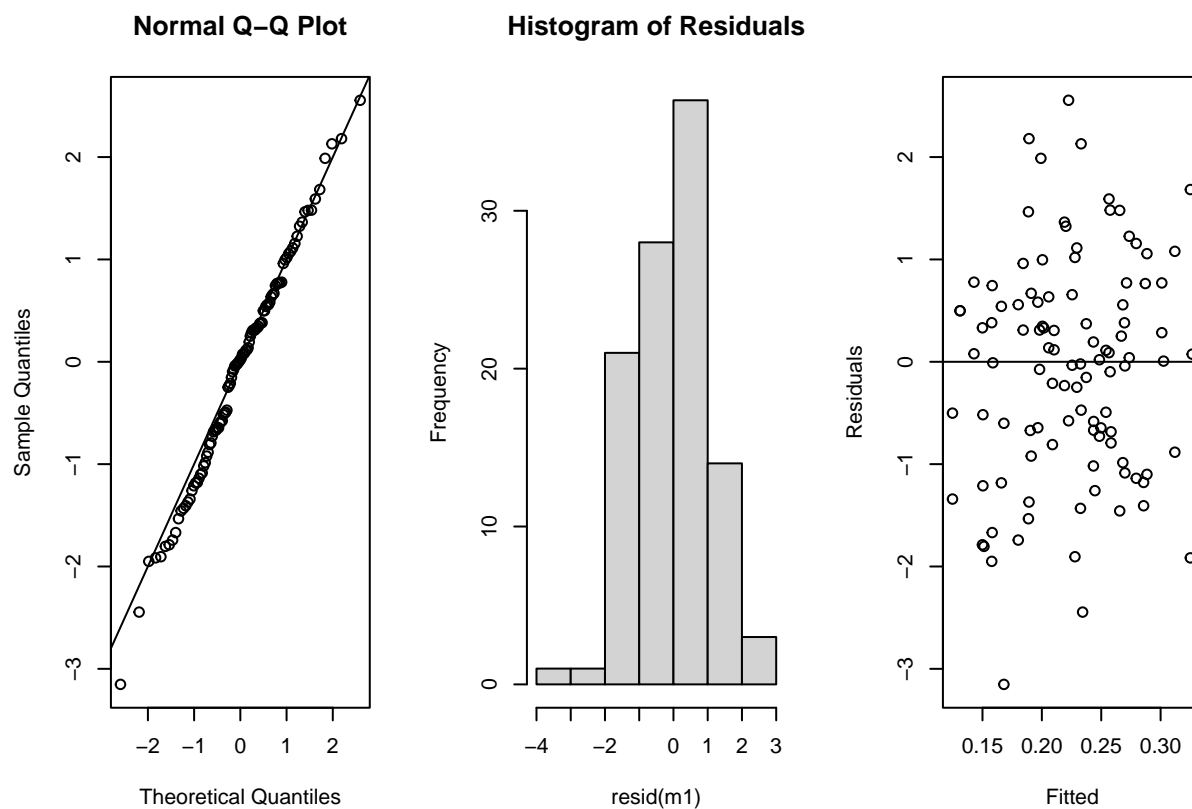
The plots above show the distribution of the residuals.

```
# Create plots of random effects
```