

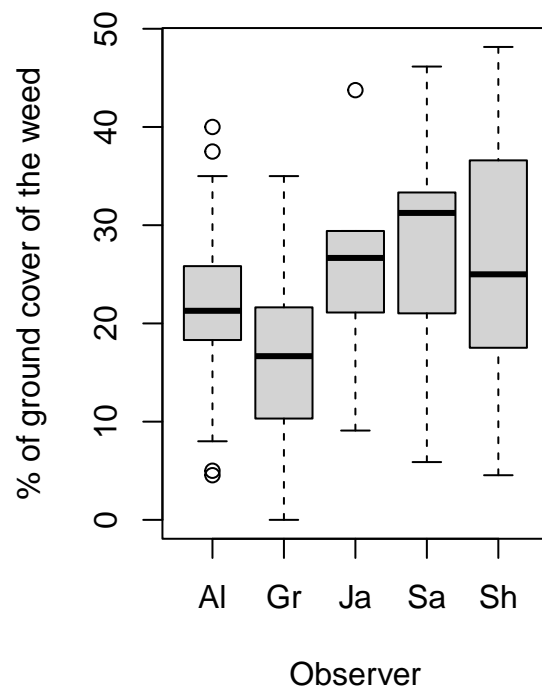
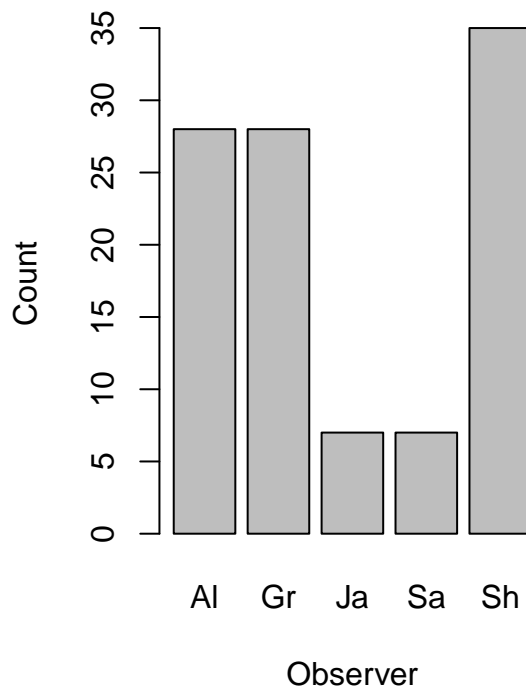
# Assignment 5 STAT 315-463: Multivariable Statistical Methods and Applications

Lisa Lu 31088272

```
# Read in data file
initobs <- read.table("initobs.csv", header = TRUE, sep = ',', na.strings = "na")
```

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

```
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")
```



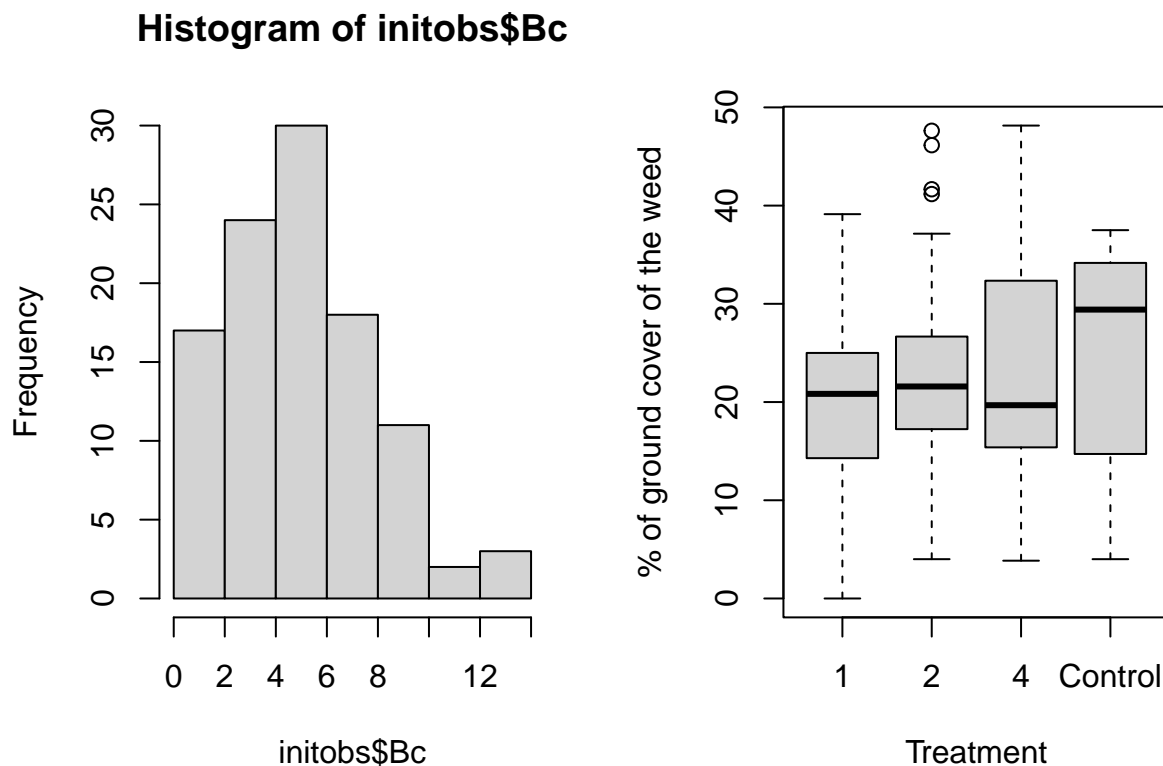
From the left plot above, we can see that the number of occurrence of each observer varies greatly. The right plot also suggested that the distribution of the percentage of ground covered by the weed differs from each observer. Therefore, the observers are the source of random variation, which should be taken into

consideration when choosing the model as it brings potential effects on the relationship between the control of the weed and the different treatments.

**b) What distribution would be appropriate for these data and why?**

The dependent variable here, namely, the occurrence of the weed (buttercup) are not continuous. In a more visualised way, from the histogram of the frequency of the buttercup shown in the left plot, we can see that it does not follow the normal distribution. Therefore it is not suitable to use normal linear regression models. Here, we would like to investigate the control of the weed by using different treatments. Each test has two potential outcomes: Has Bc or No Bc, and each test is exclusive or independent of one another. As shown in the right plot below, the data follows a discrete distribution. Therefore, binomial distribution should be considered as an appropriate way to fit the model.

```
pardef <- par()
par(mfrow = c(1,2))
hist(initobs$Bc)
plot(as.factor(initobs$Trt), initobs$Bc/initobs$Steps * 100,
     xlab="Treatment",ylab="% of ground cover of the weed")
```



**c) Fit an appropriate random effects model to these data**

```
library(lme4)
initobs$Plot <- as.factor(initobs$Plot)
m1 <- glmer(cbind(Bc, NotBc) ~ Trt + Farm + Plot + (1|Observer),
            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 + Farm + Plot + (1 | Observer)
## Data: initobs
##
##      AIC      BIC   logLik deviance df.resid
##    469.8    501.7   -222.9   445.8      93
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.0735 -0.8022  0.0589  0.5238  3.7300
##
## Random effects:
## Groups Name Variance Std.Dev.
## Observer (Intercept) 0.045 0.2121
## Number of obs: 105, groups: Observer, 5
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.57682    0.21374  -7.377 1.62e-13 ***
## Trt2          0.18289    0.19774   0.925 0.35502
## Trt4          0.49236    0.18796   2.620 0.00880 **
## TrtControl    0.49519    0.19142   2.587 0.00968 **
## FarmB         0.20066    0.20077   0.999 0.31757
## FarmF        -0.31418    0.20819  -1.509 0.13126
## FarmR        -0.39441    0.21333  -1.849 0.06448 .
## FarmS         0.05463    0.19537   0.280 0.77978
## Plot2         0.45813    0.18127   2.527 0.01149 *
## Plot3        -0.00930    0.17666  -0.053 0.95802
## Plot4         0.50351    0.18949   2.657 0.00788 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Trt2   Trt4   TrtCnt FarmB  FarmF  FarmR  FarmS  Plot2
## Trt2          -0.484
## Trt4          -0.507  0.552
## TrtControl    -0.503  0.542  0.570
## FarmB         -0.523  0.009  0.004  0.015
## FarmF         -0.490 -0.004 -0.007  0.002  0.725
## FarmR         -0.480 -0.001 -0.005 -0.005  0.709  0.681
## FarmS         -0.522 -0.002 -0.002 -0.001  0.768  0.738  0.722
## Plot2         0.006 -0.518  0.000  0.000 -0.004 -0.015 -0.010 -0.003
## Plot3         0.006  0.001 -0.446  0.001  0.001 -0.010 -0.010 -0.011  0.000
## Plot4        -0.513  0.547  0.576  0.565  0.021  0.007  0.007  0.008  0.000
##              Plot3
## Trt2
## Trt4
## TrtControl
## FarmB
## FarmF
## FarmR
## FarmS

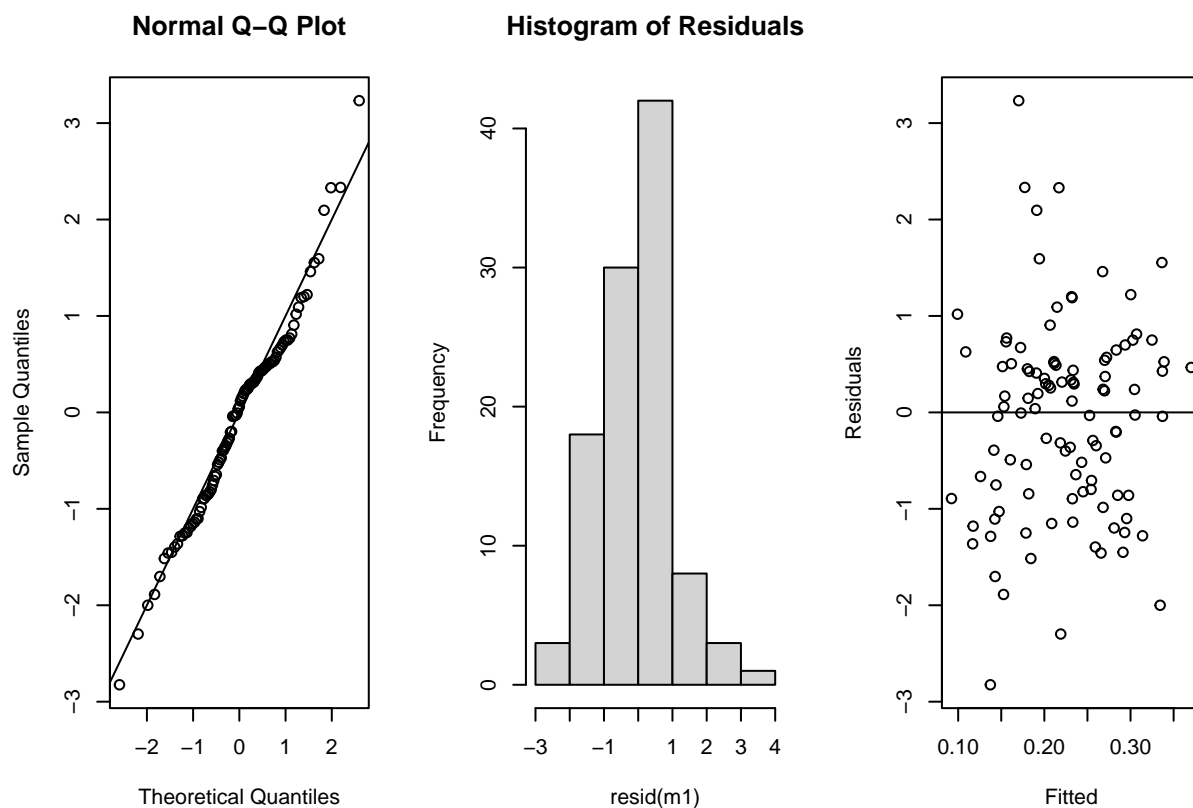
```

```
## Plot2
## Plot3
## Plot4      0.000
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 3 columns / coefficients
```

- 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)
```



# The plots above show the distribution of the residuals.

```

# Create plots of random effects
re1 <- ranef(m1)$Observer
# re2 <- ranef(m1)$Farm
# re <- data.frame(Observer = re1, Farm = re2)
# boxplot(re, col=c("lightblue", "lightpink"), main="Random Effects Boxplot",
#   # names=c("Observer", "Farm"),ylab="Random Effects")
boxplot(re1, main="Random Effects Boxplot",
  xlab="Observer",ylab="Random Effects")

```

