Exercise 2

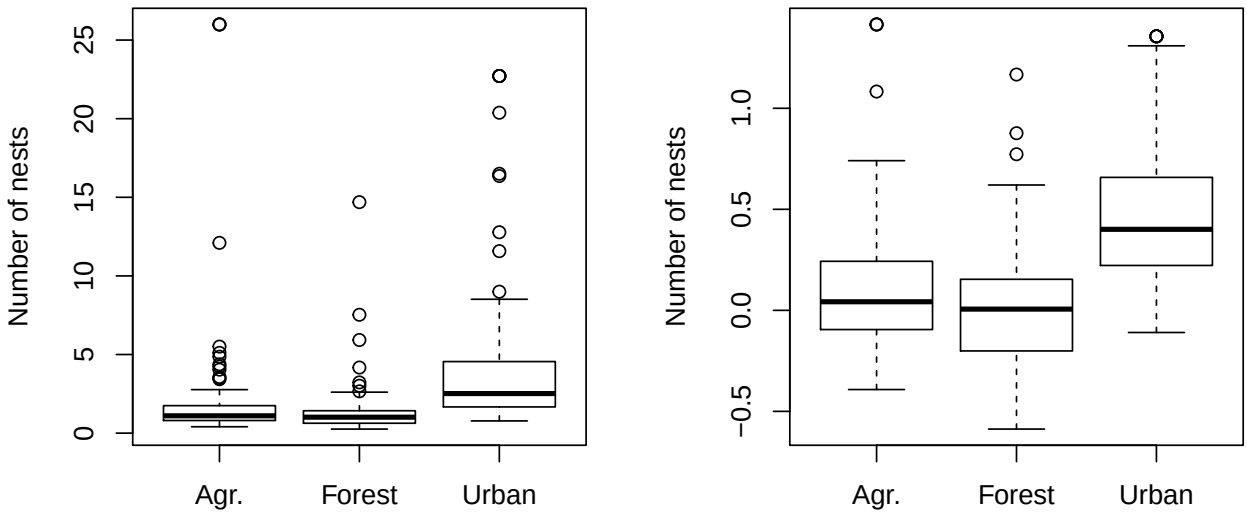
Background

We want to know what is the effect of urbanization on the reproduction of spiders. We have strong evidence to think that spiders have a higher fitness (number of offspring) in urban areas. To test this, we sample the number of spider nests in three different areas: agricultural area, forested area and urban area. The null hypothesis is that there is not difference between the three areas.

Test the hypothesis

First, navigate to your working directory with *setwd()* and load the dataset called *spider-nest.csv* into the environment using *read.csv()*. Familiarize with the dataset and display the boxplot of the count of number of nests per area. This can be done using the function *boxplot(y ~ x)*.

```
setwd('/home/GIT/BEHAVIOURAL-BIOLOGY-2019/non-par_ANOVA')
dataset <- read.csv('spider-nest.csv')
par(mfrow = c(1, 2)) # 1 row, 2 columns in the figure
boxplot(Number.of.nests ~ Area, data = dataset, ylab = 'Number of nests')
boxplot(log10(Number.of.nests) ~ Area, data = dataset, ylab = 'Number of nests')
```

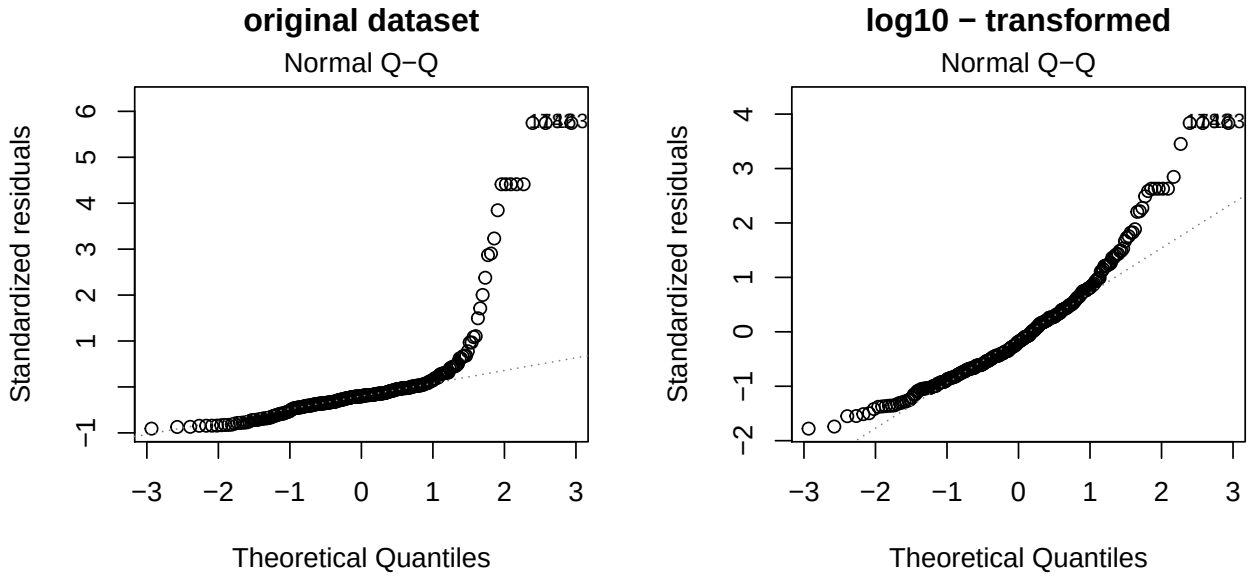


Counts of objects in an area are rarely normally distributed variables. To check if this is the case, we create a linear model and plot its residuals.

```
par(mfrow = c(1, 2))

model <- lm(Number.of.nests ~ Area, data = dataset)
plot(model, which = 2, main = 'original dataset')

model <- lm(log10(Number.of.nests) ~ Area, data = dataset)
plot(model, which = 2, main = 'log10 - transformed')
```



Question 1: are these data normally distributed?

Question 2: using the Kruskal-Wallis test, accept or reject the null hypothesis: there is not difference in the number of nests between areas.

```
kruskal.test(dataset$Number.of.nests, dataset$Area)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  dataset$Number.of.nests and dataset$Area
## Kruskal-Wallis chi-squared = 91.152, df = 2, p-value < 2.2e-16
```

Question 3: using the rank-sum test test, find the areas that differ in number of nests with an overall significance level of $\bar{\alpha} = 0.05$ (use Bonferroni correction). Which ones are different? In R the rank-sum test is implemented in the *wilcox.test()* function, which contains the parameters *paired* that determines if the rank-sum test is performed ($paired = F$), or the signed-rank test is performed ($paired = T$).

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  urban and forest
## W = 8645, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  urban and agr
## W = 7942, p-value = 6.609e-13
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  agr and forest
## W = 5973, p-value = 0.01749
## alternative hypothesis: true location shift is not equal to 0
```