

```
# IRDR04 Research Tools Module
# Week 8 Tutorial
# 27/11/18 9-11am @ Birkbeck, Malet Street 414/415 Public Cluster
# Tutor: Dr Alison Fairbrass, UCL Centre for Biodiversity & Environment Research, alison.fairbrass@gmail.com
```

```
# # # Exercises and Solutions # # #
```

```
# Exercise 1
# Create a high quality plot using the datafile: Nelson.csv
```

```
# Solution
# Students can plot the data in which ever way they like, but I want to see them using the functionality of R to make
# the plot visually appealing.
```

```
# Exercise 2
# Load a new dataset mussel.csv
# The study investigated abundance-area effects for invertebrates living in mussel beds in intertidal areas
# 25 mussel beds
# response = number of invertebrates (INDIV)
# Explanatory = the area of each clump (AREA)
# additional possible response - Species richness of invertebrates (SPECIES)
# We're going to look at species richness
```

```
# a) Plot the data and assess whether a linear regression is appropriate.
```

```
# Solution
# load data file
Mussel <- read.csv(file.choose())
```

```
# Look at it
str(Mussel)
head(Mussel)
```

```
scatterplot(SPECIES ~ AREA, data = Mussel)
# This indicates that the data are not normally distributed (especially AREA)
```

```
# b) Fit a linear model of AREA against SPECIES. What do the results of the model tell you about
# the relationship between area and species richness?
```

```
# Solution
mussel.lm <- lm(SPECIES ~ AREA, data = Mussel)
```

```
summary(mussel.lm)
```

```
# The results suggest that there is a significant positive correlation between the variables.
# The adjusted R-square suggested that there is some unexplained variation in the model.
# We should try adding more variables to the model to account for this variation.
```

```
# c) Apply a log transformation to the AREA variable (as it is not normally distributed)
# and re-run the model.
```

```
# Solution
mussel.lml <- lm(SPECIES ~ log10(AREA), data = Mussel)
```

```
summary(mussel.lml)
```

```
# d) Use appropriate plots to check which model fits the assumptions of linear regression better.
# Explain your answer.
```

```
# Solution
# Check 1st model
plot(fitted(mussel.lm), residuals(mussel.lm))
hist(residuals(mussel.lm))
qqnorm(residuals(mussel.lm))
qqline(residuals(mussel.lm))
```

```
# Check 2nd model
plot(fitted(mussel.lml), residuals(mussel.lml))
hist(residuals(mussel.lml))
qqnorm(residuals(mussel.lml))
qqline(residuals(mussel.lml))
```

```
# The second model is better as there is no pattern in the residual vs fitted plot, and the residuals
# in the QQ plot are more closely fit to the line.
```