Stats*3510 - Assignment 2
0679576 - Graham Eckel

Q2
```
library(ggplot2)
dir = "C:\\Users\\...\\Stat3510 - Environmental Risk Analysis\\Assignment 2\\"
file1 = "Bird_Extinction_Data.csv"
dfBirdExtinct = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')
dfBirdExtinct$logtime = log(dfBirdExtinct$time)
```

a)
```
p = ggplot(dfBirdExtinct, aes(pairs, logtime, colour=size, shape=migration,  xlab("Nesting
Pairs"))) + geom_point()
p + labs(x = "Nesting Pairs", y = "Log of Average Time to Extinction", title = "Scatterplot of Log
of Average Time to Extinction and Nesting Pairs of Birds")
```
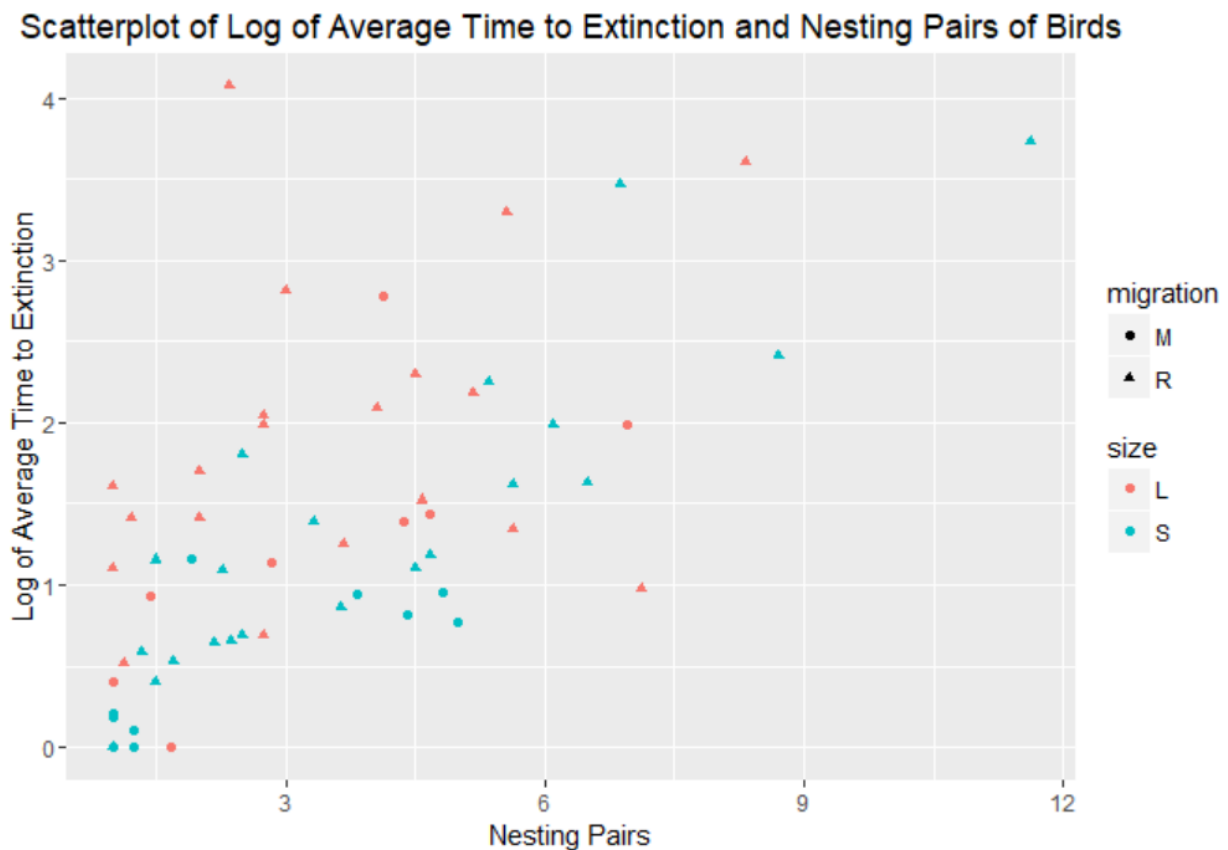
b)

bird.modelAdd = lm(logtime~pairs+size+migration, data = dfBirdExtinct)
summary(bird.modelAdd)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.43087    0.20706   2.081 0.041870 *
pairs         0.26501    0.03679   7.203 1.33e-09 ***
sizeS        -0.65220    0.16667  -3.913 0.000242 ***
migrationR    0.50417    0.18263   2.761 0.007712 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6524 on 58 degrees of freedom
Multiple R-squared:  0.5982,    Adjusted R-squared:  0.5775
F-statistic: 28.79 on 3 and 58 DF,  p-value: 1.577e-11
```

bi)

With p-values of 0.00000000133, 0.000242 and 0.007712, there is evidence that the variables nesting pairs, small size and resident migration and are all significant in estimating the time to extinction, given they are all in the model.

bii)

size = small (S) = 1   migration = resident (R) = 1
size = large (L) = 0   migration = migratory (M) = 0

(1) uhat(y|size=1, migration = 0)   = 0.43087 + 0.2651(B1) - 0.65220(B2)
                                     = -0.22133 + 0.2651(B1)

(2) uhat(y|size=1, migration = 1)   = 0.43087 + 0.2651(B1) - 0.65220(B2) + 0.50417(B3)
                                     = 0.28284 + 0.2651(B1)

(3) uhat(y|size=0, migration = 0)   = 0.43087 + 0.2651(B1)

(4) uhat(y|size=0, migration = 1)   = 0.43087 + 0.2651(B1) + 0.50417(B3)
                                     = 0.93504 + 0.2651(B1)

B0hat (estimated as 0.43087) is the intercept for large, migratory birds and there is evidence (p = 0.04187) that it is different from 0 in this population. This estimate represents the average log time to extinction for large birds that migrate and do not have a nesting pair.

B1hat is coefficient of nesting pairs. This is the effect of every added nesting pair on all birds and there is strong evidence (p = 0.00000000133) that this is different from 0 in the population

B2hat is the difference in estimated log time to extinction between a large and small bird, after adjusting for their migratory habits and given they have the same nesting pairs

B3hat is the difference in estimated log time to extinction between resident and migratory birds, after adjusting for their size and give they have the same nesting pairs.

biii) if I were to hand sketch this model, it would look like 4 parallel lines with y intercepts of -0.22133, 0.28284, 0.43087 and 0.93504 and slopes of 0.2651.

c)
bird.modelInteract = lm(logtime~pairs+size+migration+pairs*size, data = dfBirdExtinct)
summary(bird.modelInteract)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.67221    0.27762   2.421  0.01867 *
pairs        0.20059    0.06173   3.250  0.00194 **
sizeS       -0.99372    0.31135  -3.192  0.00230 **
migrationR   0.48071    0.18247   2.634  0.01083 *
pairs:sizeS  0.09897    0.07638   1.296  0.20032
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6486 on 57 degrees of freedom
Multiple R-squared:  0.6097,     Adjusted R-squared:  0.5823
F-statistic: 22.26 on 4 and 57 DF,  p-value: 4.153e-11
```

ci) B4hat (estimated as 0.09897) is the difference in the intercept for small birds with one or more nesting pairs after adjusting for migratory behaviour. It also gives the difference in slopes between models.
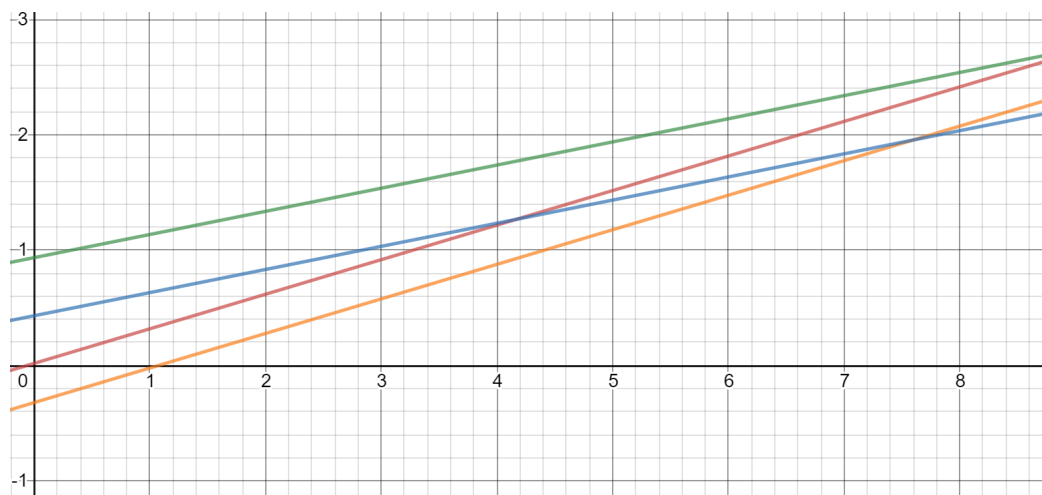
cii)

(1) uhat(y|size=1, migration = 0)　　= 0.67221 + 0.20059(B1) - 0.99372(B2) + 0.09897(B1*B2)
　　　　　　　　　　　　　　　　　　= -0.32151+0.29956(B1)

(2) uhat(y|size=1, migration = 1)　　= 0.43087 + 0.20059(B1) - 0.99372(B2) + 0.50417(B3)
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　+ 0.09897(B1*B2)

　　　　　　　　　　　　　　　　　　= 0.018266+ 0.29956(B1)

(3) uhat(y|size=0, migration = 0)　　= 0.43087 + 0.20059(B1)

(4) uhat(y|size=0, migration = 1)　　= 0.43087 + 0.20059(B1) + 0.50417(B3)
　　　　　　　　　　　　　　　　　　= 0.93504 + 0.20059(B1)

It's a bit challenging to see, but green (1) and blue (2) are parallel, red (3) and orange (4) are parallel, and (1) and (2) intersect (3) and (4).

ciii)
The interaction effect between nesting pairs is not significant (p = 0.20032) and there is not enough evidence to believe the difference, after adjusting for the other variables, is not 0 in the population.

civ)
anova(bird.modelAdd, bird.modelInteract)

```
Analysis of Variance Table

Model 1: logtime ~ pairs + size + migration
Model 2: logtime ~ pairs + size + migration + pairs * size
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     58 24.687
2     57 23.981  1   0.70626 1.6787 0.2003
>
```

Ho: The reduced model, without adjusting for an interaction between bird size and number of nesting pairs is adequate
Ha: The full model, adjusting for the interaction is necessary

There is insufficient evidence (p = 0.2003) that adjusting for an interaction between size and number of nesting pairs is necessary in a model for estimating log time until extinction of birds after adjusting for size and migration habits.

d)
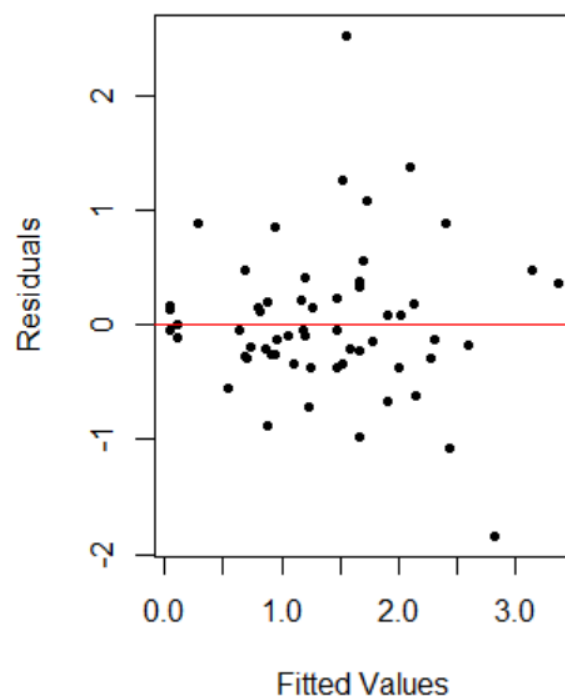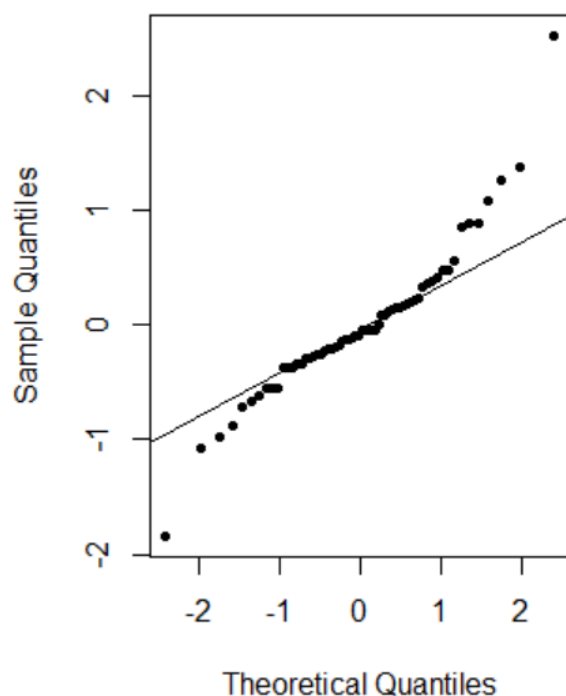
The assumption of linearity appears to hold as the points are randomly distributed in the Residuals plot. The assumption of normality does not appear to hold as there are points quite far away from the straight line in the QQ plot. The assumption of constant variance appears to hold as the points are evenly distributed across the midline in the Residuals plot.