homework3

Chimdi Chikezie

4/26/2023

```
library(Sleuth3)
library(ggplot2)
library(Rmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: plyr
```

```
library(graphics)
```

###Problem 1: Data taken from Exercise 10.27 Factors Affecting Extinction. The data are measurements on breeding pairs of land-bird species collected from 16 islands around Britain over the course of several decades. For each species, the data set contains an average time of extinction on those islands where it appeared (this is actually the reciprocal of the average of $1/T$, where T is the length of time the species remained on the island, and $1/T$ is taken to be zero if the species did not become extinct on the island, ignore this explanation if it confuses you); the average number of nesting pairs (the average, over all islands where the birds appeared, of the number of nesting pairs per year); the size of the species (categorized as large or small); and the migratory status of the species (migrant or resident). (Data from S. L. Pimm, H. L. Jones, and J. Diamond, "On the Risk of Extinction, American Naturalist 132 (1988): 757-85.) It is expected that species with larger numbers of nesting pairs will tend to remain longer before becoming extinct. So, we consider average time of extinction as the response variable and the average number of nesting pairs, the size of the species and the migratory status of the species as the explanatory variables. (Each part of question carries 4 points

```
#?ex1027
dataset <- ex1027
#dataset
```

a) Consider the parallel lines model for multiple linear regression of average time of extinction against average number of nesting pairs, the size of the species and the migratory status of the species (since, the last two variables are categorical variables, you will need to convert them into corresponding indicator variables). Test the hypothesis that size of the species does not have any effect on the average time of extinction (format of reporting hypothesis testing - write the null and alternate hypothesis in terms of model parameters and also in sentence, perform the hypothesis testing, report p-value and state the conclusion in plain english).

```
dataset$Size <- as.factor(dataset$Size)
```

```
lin_mod <- lm(Time ~ Pairs + Size + Status, data = dataset)
summary(lin_mod)
```

```
##
## Call:
## lm(formula = Time ~ Pairs + Size + Status, data = dataset)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -15.424  -4.259  -1.586   0.774  50.905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6296     2.9117  -0.216  0.82956
## Pairs         2.1339     0.5173   4.125  0.00012 ***
## SizeS        -3.9269     2.3435  -1.676  0.09919 .
## StatusR       3.5300     2.5679   1.375  0.17453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.173 on 58 degrees of freedom
## Multiple R-squared:  0.3007, Adjusted R-squared:  0.2645
## F-statistic: 8.312 on 3 and 58 DF,  p-value: 0.0001097
```

$H_0 : \mu_L = \mu_S$

$H_A : \mu_L \neq \mu_S$

Where L = SizeLarge and S = SizeSmall

The parameter that measures the difference between the intercepts of the size of the species is $\beta_2 = -3.9269$. The p-value(0.09919) associated with the t-test is greater than $\alpha = 0.05$, which means that $\beta_2 = 0$. We fail to reject the null hypothesis that $\beta_2 = 0$ and conclude that size of the species does not have any effect on the average time of extinction.

b) Test the hypothesis that the migratory species of small size does not have a different average time of extinction than resident species of small size with the same average number of nesting pairs.

$H_0 : \mu_M = \mu_R$

$H_A : \mu_M \neq \mu_R$

Where M = StatusMigrant and R = StatusResident

The parameter that measures the difference between the intercepts of the status of the species is $\beta_2 = 3.5300$. The p-value(0.17453) associated with the t-test is greater than $\alpha = 0.05$, which means that $\beta_3 = 0$. We fail to reject the null hypothesis that $\beta_3 = 0$ and conclude that the migratory species of small size does not have a different average time of extinction than resident species of small size with the same average number of nesting pairs.

(c) Find a 90% confidence interval for the parameter tested in part (a). Write a sentence explaining the confidence interval. Find a 99% confidence interval for the parameter tested in part (b). Write a sentence explaining the confidence interval.

```
'90% confidence interval for the parameter tested in part (a)'
```

```
## [1] "90% confidence interval for the parameter tested in part (a)"
```

```
confint(lin_mod, level = 0.90)
```

```
##                    5 %         95 %
## (Intercept) -5.4967092   4.23743883
## Pairs        1.2691946   2.99855219
## SizeS       -7.8441829  -0.00960805
## StatusR     -0.7624134   7.82238626
```

With 90% confidence, the average time of extinction of the SizeS birds is between -7.8441829 and -0.00960805 less than the time of extinction of the SizeL birds.

```
'99% confidence interval for the parameter tested in part (b)'
```

```
## [1] "99% confidence interval for the parameter tested in part (b)"
```

```
confint(lin_mod, level = 0.99)
```

```
##                      0.5 %      99.5 %
## (Intercept)   -8.3843499    7.125079
## Pairs          0.7561796    3.511567
## SizeS        -10.1683141    2.314523
## StatusR       -3.3090992   10.369072
```

With 99% confidence, the difference in average time of extinction of migratory species and resident species of the same size and average number of nesting pairs is between -3.3090992 and 10.369072.

   (d) Consider the separate lines model for multiple linear regression of average time of extinction against average number of nesting pairs, the size of the species and the migratory status of the species, by considering all three pairwise interaction terms and also the three-way interaction term. Compare this full interaction model with the model that has only two interaction terms - the interaction term between average number of nesting pairs and size of species and the interaction term between average number of nesting pairs and migratory status. Do this comparison by using an ANOVA test.

```
## Continuous Intensity Variable
complex <- lm(Time ~ Pairs * Size * Status, data = dataset)
simple <- lm(Time ~ Pairs + Size + Status + (Pairs*Size) + (Pairs*Status), data = dataset)
## Comparing Models
anova(complex, simple)
```

```
## Analysis of Variance Table
##
## Model 1: Time ~ Pairs * Size * Status
## Model 2: Time ~ Pairs + Size + Status + (Pairs * Size) + (Pairs * Status)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     54 4607.7
## 2     56 4686.3 -2   -78.584 0.4605 0.6334
```

The p-value(0.6334) from the comparison is greater than $\alpha = 0.05$ which means the comparison is not statistically significant. Therefore, we conclude that the simpler model is better to explain the data than the complex model.

   (e) Consider the separate lines model that has only two interaction terms - the interaction term between average number of nesting pairs and size of species and the interaction term between average number of nesting pairs and migratory status only. Test the hypothesis that interaction terms in the model are not needed, that is, test the null hypothesis that the difference between interaction model and parallel lines model of part (a) is not significant.

```
simple <- lm(Time ~ Pairs + Size + Status + (Pairs*Size) + (Pairs*Status), data = dataset)
summary(simple)
```

```
##
## Call:
## lm(formula = Time ~ Pairs + Size + Status + (Pairs * Size) +
##       (Pairs * Status), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.499  -3.774  -1.207   1.473  50.724
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.1102     4.8673   1.050    0.298
## Pairs           0.3228     1.3083   0.247    0.806
## SizeS          -6.4163     4.3918  -1.461    0.150
## StatusR        -1.9762     4.6870  -0.422    0.675
## Pairs:SizeS     0.5781     1.0839   0.533    0.596
## Pairs:StatusR   1.7888     1.3146   1.361    0.179
##
## Residual standard error: 9.148 on 56 degrees of freedom
## Multiple R-squared:  0.3285, Adjusted R-squared:  0.2686
## F-statistic:  5.48 on 5 and 56 DF,  p-value: 0.0003564
```

$H_0 : interaction terms = 0$

$H_A : interaction terms \neq 0$

The p-values of the interaction terms Pairs:SizeS(0.596) and Pairs:StatusR(0.179) are greater than $\alpha = 0.05$. This makes them not statistically significant. Therefore, we fail top reject the null hypothesis that the interaction terms in the model are equal to zero, which means they are not needed.

(f) Give the ANOVA table for the model in part (e). Explain the table and test the significance of the model using ANOVA F-test.

```
anova(simple)
```

```
## Analysis of Variance Table
##
## Response: Time
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Pairs         1 1688.4 1688.40 20.1760 3.581e-05 ***
## Size          1  251.0  251.01  2.9995   0.08879 .
## Status        1  159.0  159.02  1.9002   0.17353
## Pairs:Size    1   39.6   39.56  0.4727   0.49459
## Pairs:Status  1  155.0  154.95  1.8517   0.17904
## Residuals    56 4686.3   83.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this model, Pairs has p-value less than $\alpha = 0.05$, indicating a significant relationship between average time of extinction and average number of nesting Pairs but the other parameters are not significant because the have p-values greater than $\alpha = 0.05$. This indicates that only average number of nesting Pairs has a significant effect on average time of extinction, while the other parameters do not have a significant impact.

This means that the model is significant, and we can reject the null hypothesis that there is no relationship between the dependent and at least one independent variables.