Multiple linear regression

October 16, 2020

Answers for this lab must be submitted before **October 28th at 5pm on Moodle**. In your answer for each question, please include the R code used (if applicable) and the results obtained.

1. Sablefish catches in Alaska

The file sablefish.csv contains data from Kimura (1988) on catches of sablefish per unit effort in four locations in Alaska for each of the six years between 1978 and 1983.

```
sable <- read.csv("sablefish.csv")
str(sable)
```

```
## 'data.frame':    24 obs. of  3 variables:
##  $ year    : int  1978 1978 1978 1978 1979 1979 1979 1979 1980 1980 ...
##  $ location: chr  "Shumagin" "Chirikof" "Kodiak" "Yakutat" ...
##  $ catch   : num  0.236 0.204 0.241 0.232 0.14 0.202 0.228 0.268 0.286 0.275 ...
```

a) Convert the year and location to factors in R, then perform an ANOVA to determine if abundance varies significantly from year to year ($\alpha = 0.05$). From the diagnostic graphs, verify that the assumptions of the ANOVA model are met.

Note: We suppose here that the effects are additive. Also, since there is only one measurement for each combination of a year and a location, it is not possible to estimate an interaction.

b) Re-analyze the model in (a) with the linear regression function `lm`. Use the appropriate contrasts to determine the mean catch and the deviation from this mean for each year.

c) According to the results in (b), which location has the greatest mean catch?

d) Using the *emmeans* package, illustrate the estimated mean catch for each year with confidence intervals. Then, using a multiple comparisons test, indicate between which years the catch varies significantly.

2. Metabolism of a fish according to salinity

The dataset sardinella.csv comes from a study by Wohlschlag (1957), "Differences in metabolic rates of migratory and resident freshwater forms of an Arctic Whitefish". It contains weight (*log_weight*) and oxygen consumption (*log_O2*) measurements for individuals of *Coregonus sardinella* caught in freshwater or marine environments.

```
sardinella <- read.csv("sardinella.csv")
str(sardinella)
```

```
## 'data.frame':    22 obs. of  3 variables:
##  $ environment: chr  "marine" "marine" "marine" "marine" ...
##  $ log_O2     : num  1.59 1.4 1.47 1.66 1.55 ...
##  $ log_weight : num  2.5 2.04 2.15 2.35 2.24 ...
```

a) Estimate the additive effects of environment and weight on the oxygen consumption of this fish. How do you interpret each of the parameters of the model?

b) Repeat the model in (a) with a standardized version of the predictor *log_weight* (*norm_weight*). What is the interpretation of the coefficients now?

c) Repeat the model in (b) by adding the interaction between the weight (standardized) and the environment. Is this interaction meaningful? What is the interpretation of the coefficients?