Homework 04

Adam Guerra

5/17/23

Problem 1

1. Null and Alternative Hypothesis

Mathematical

- $H_0 : \beta_1 = 0$
- $H_A : \beta_1 \neq 0$

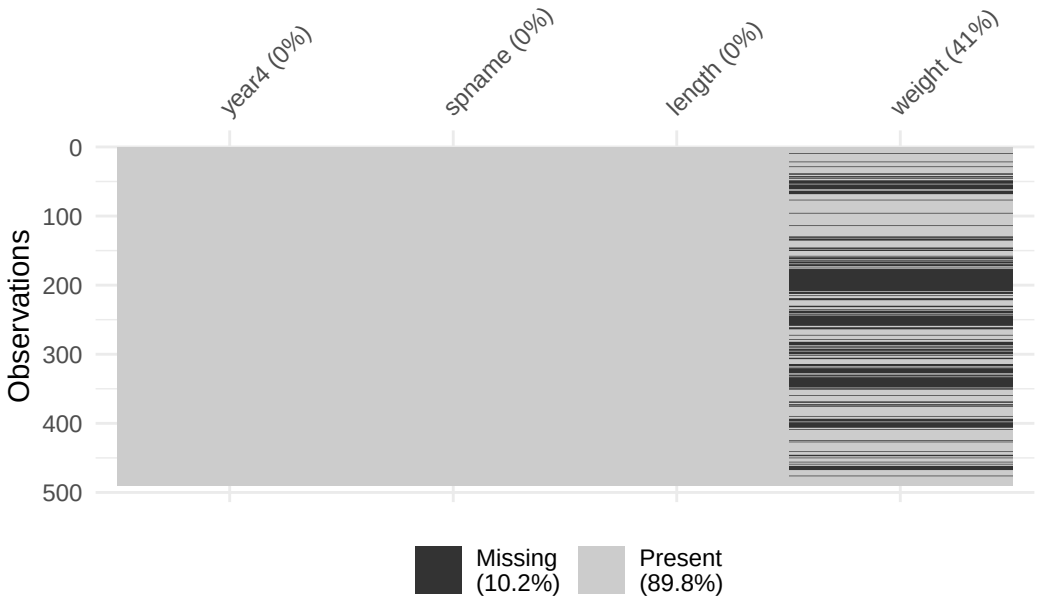
Biological

- H_0 : Fish length is not a significant predictor of fish weight for trout perch.
- H_A : Fish length is a significant predictor of fish weight for trout perch.

2. Visualize the missing data

```
#clean and select data
fish_data <- sqldf("SELECT year4, spname, length, weight
                    FROM fish_raw_data
                    WHERE spname = 'TROUTPERCH'")

#visualize missing data
vis_miss(fish_data)
```



As shown in the figure above, 41% of the weight category has NAs. This is going to limit how many observations I have to use in the linear model, reducing its statistical power.

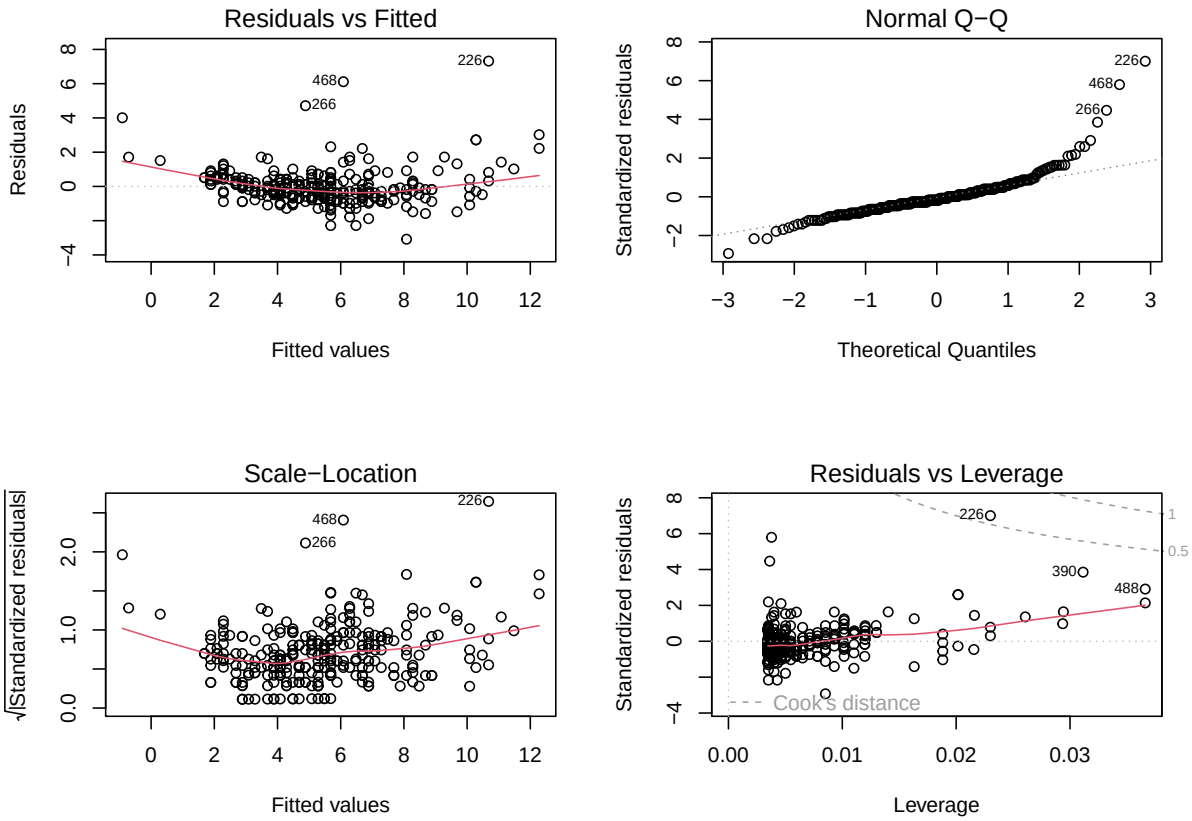
3. Run test

```
#create linear model
fish_lm <- lm(weight ~ length, data = fish_data)

#get residuals
fish_res <- fish_lm$residuals
```

4. Visually check assumptions

```
#diagnostic plots
par(mfrow = c(2,2))
plot(fish_lm)
```



- Residuals vs Fitted: Shows residuals and fitted line to visualize constant variance. Dots appear to be evenly and randomly distributed around the line.
- QQ: Shows both data sets against one another. Data appears to be normally distributed.
- Scale Location: Similar to residuals vs fitted, showing homoscedasticity of variance. Data appears to be evenly and randomly distributed around the line.
- Residuals vs Leverage: Shows which data points are influential in the model. There are a few outliers identified that could be influential.

Repo Link