

SOUTHERN METHODIST UNIVERSITY  
MSDS 6371(401)

---

# Kaggle Project

---

Michael J Wolfe  
Sandesh Ojha  
Carl Walenciak  
Travis Daun

February 14, 2019

## 1. INTRODUCTION

## 2. DATA DESCRIPTION

## 3. ANALYSIS QUESTION 1

### 3.1. Restatement of Problem

### 3.2. Build and Fit the Model

### 3.3. Checking Assumptions

#### 3.3.1. Residual Plots

#### 3.3.2. Influential point analysis (Cook's $D$ and Leverage)

#### 3.3.3. Make sure to address each assumption

### 3.4. Model Metrics

#### 3.4.1. Adj $R^2$

#### 3.4.2. Internal CV Press

### 3.5. Parameters

#### 3.5.1. Estimates

#### 3.5.2. Interpretation

#### 3.5.3. Confidence Intervals

### 3.6. Conclusion

## 4. ANALYSIS QUESTION 2

### 4.1. Restatement of Problem

### 4.2. Model Selection

#### 4.2.1. Stepwise

#### 4.2.2. Forward

#### 4.2.3. Backward

#### 4.2.4. Custom

### 4.3. Checking Assumptions

#### 4.3.1. Residual Plots

#### 4.3.2. Influential point analysis (Cook's $D$ and Leverage)

#### 4.3.3. Make sure to address each assumption

### 4.4. Comparing Competing Models

#### 4.4.1. Adj $R^2$

#### 4.4.2. Internal CV Press

#### 4.4.3. Kaggle Score

### 4.5. Conclusion

Predictive Models	Adjusted R <sup>2</sup>	CV Press	Kaggle Score
Forward	XX	XX	XX
Backward	XX	XX	XX
Stepwise	XX	XX	XX
CUSTOM	XX	XX	XX

Table 4.1: Analysis Results

## A. SOURCE CODE FOR ANALYSIS 1

Listing 1: Analysis 1 SAS Code.

```

1  /* Dummy Code File for Group Project */
2
3  data crabs;
4      infile '/folders/myfolders/MSDS6371/HW12/Datasets/Crab17.csv'
5      missover delimiter=', ' firstobs=2;
6      input Force Height Species $;
7      logForce = log(Force);
8      logHeight = log(Height);
9      sqrtForce = sqrt(Force);
10     sqrtHeight = sqrt(Height);
11     invForce = 1/Force;
12     invHeight = 1/Height;
13 run;
14
15 PROC REG DATA=crabs;
16     model Force = Height;
17     model logForce = Height;
18     model Force = logHeight;
19     model logForce = logHeight;
20     model sqrtForce = Height;
21     model Force = sqrtHeight;
22     model sqrtForce = sqrtHeight;
23     model invForce = Height;
24     model Force = invHeight;
25     model invForce = invHeight;
26 RUN;
```

## B. SOURCE CODE FOR ANALYSIS 2

Listing 2: Analysis 2 SAS Code.

```
1  /* Dummy Code File for Group Project */
2
3  data crabs;
4      infile '/folders/myfolders/MSDS6371/HW12/Datasets/Crab17.csv'
5      missover delimiter=', ' firstobs=2;
6      input Force Height Species $;
7      logForce = log(Force);
8      logHeight = log(Height);
9      sqrtForce = sqrt(Force);
10     sqrtHeight = sqrt(Height);
11     invForce = 1/Force;
12     invHeight = 1/Height;
13 run;
14
15 PROC REG DATA=crabs;
16     model Force = Height;
17     model logForce = Height;
18     model Force = logHeight;
19     model logForce = logHeight;
20     model sqrtForce = Height;
21     model Force = sqrtHeight;
22     model sqrtForce = sqrtHeight;
23     model invForce = Height;
24     model Force = invHeight;
25     model invForce = invHeight;
26 RUN;
```