PART B

In this part, please upload the provided data set into R. The dataset you upload depends on your discussion section. The dataset is in "FinalData.csv"

**You are expected to turn in a .R** file that includes all commands you used to prepare your answers for Part B. (You need not include any calculations you may have performed for part A.)

The data are from a study you met on the midterm and are based on a random sample of "Coconut" crabs found in Japan that are famous for their extreme pinching force.

The variable *size* is the sum of the variables *ThoraxLength, ClawLength, and ClawHeight*.

1. Fit a basic linear model using only *size*, *Weight* and *Sex* to predict pinching force. Do not do any transformations or higher order terms.

a) (2) Write the equation of the model: *Predicted\_PinchingForce*=

*Predicted\_PinchingForce*= -14.52549+0.04887\*Weight -0.01597\*Sex + 0.27448\*size

b) (3) Comment on the model validity with respect to these three conditions. Type the word "is" or "isn't" and then give your reason.

Linear trend condition [is or isn't?] satisfied

because:

Constant Variance condition [is or isn't?] satisfied

because:

Normal distribution of errors condition [is or isn't?] satisfied

because:

The residual plot shows a non-linear trend.

The variance is not constant since the residual plot shows a fan shape where variance increasese

2. (2) Create an Inverse Response Plot. What transformation of *PinchingForce* provides the lowest residual sums of squares?

3. (2) What transformation of *PinchingForce* is suggested by the Box-Cox transformation?

4. (3) Fit the model using the transformation for *PinchingForce* based on the Box-Cox power transform (using the "Rounded" power). Which model do you think is better, in terms of model validity: the "basic" model in question1 or this model? Explain.

5. (3) At the midterm, we found that the pinching force for male crabs was greater than for female crabs. Explain why this is not the case with the current model. (Hint: note that male crabs tend to be bigger and heavier than females.)

## t-test pinching force male and female

B+b\*sex

In midterm we simple regression

In the midterm, we applied a t-test on pinching for on male and female and this is a simple regression. However, here we are doing a multiple regression and it is obvious that gender is relatively insignificant but size is the most significant factor.

6. (2) Give the variance inflation factors for each variable for the transformed model from question 4.:

Sex

Weight

Size

7.(2) What do these values for vif tell us in this context?

8. Perform best subsets regression, forward stepwise, and backward stepwise to develop the "best" model, using BIC as a criteria. Use your transformed version of PinchingForce. Include these predictors to start: Weight, ThoraxLength, Sex, ClawLength, ClawHeight, ClawWeight. Note that you may get three different models from each of these three approaches. Choose the one with the lowest BIC. Be sure to state the BIC value for your choice. Use this model to answer these questions:

a) (2) Give the equation for the final model you chose:

b) (2) BIC for final model:

c) (2) Suppose we had just caught a coconut crab with these measurements:

ThoraxLength: 52

Weight: 615

Sex: Male

ClawLength: 67

ClawHeight: 26

ClawWeight: 34

Predict it's pinching force at a 95% level (give the appropriate interval)

9). Consider the output below:

> summary(mfull)

Call:

lm(formula = sqrt(PinchingForce) ~ Weight + Sex + ClawLength +

ClawHeight + ClawWeight, data = crabs2)

Residuals:

Min 1Q Median 3Q Max

-0.77961 -0.39837 -0.01247 0.28257 1.16845

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.027e-01 8.733e-01 -0.576 0.572

Weight 8.123e-05 8.164e-04 0.100 0.922

SexM 6.776e-02 3.194e-01 0.212 0.834

ClawLength 6.365e-02 5.397e-02 1.179 0.255

ClawHeight 5.816e-02 7.883e-02 0.738 0.471

ClawWeight 4.621e-02 1.396e-01 0.331 0.745

Residual standard error: 0.607 on 17 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.9736, Adjusted R-squared: 0.9659

F-statistic: 125.5 on 5 and 17 DF, p-value: 8.375e-13

a) (2) What null and alternative hypotheses does the F-statistic test?

b) (2) What do you conclude based on the p-value (using a significance level of 0.05)?

c) (3) Extra Credit: What's going on here?

#we are interpret all the value, contridiction going