**UNIT 12 HW**

1. As part of a study of the effects of predatory intertidal crab species on snail populations, researchers measured the mean closing forces and the propodus heights of the claws on several crabs of three species. The crab data will be in your files repository.

* Use alpha = 0.05, where necessary.
* Use SAS and provide relevant code and output.

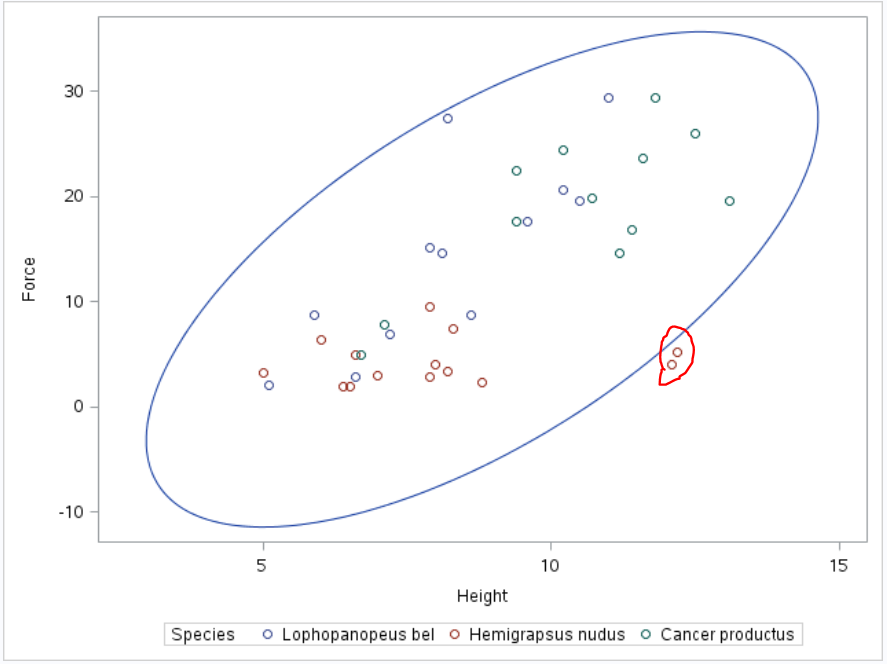
1. Step 1: Use the code from Dr. McGee’s lecture to plot a scatter plot of claw closing force (response variable) versus propodus height (explanatory variable), with different plotting symbols (or colors) to distinguish the three different crab species. Judging from an initial visual assessment of the scatterplots, you may apply a transformation and replot in this step. If a transformation is necessary, you only need to provide the scatterplot for the most visually satisfying transformation for now (but still provide a scatterplot of original data). You will formally assess the fit of the model in Step 4.

***Original Data***

**proc** **sgplot** data=work.crabsraw;

scatter x=Height y=Force / group=Species;

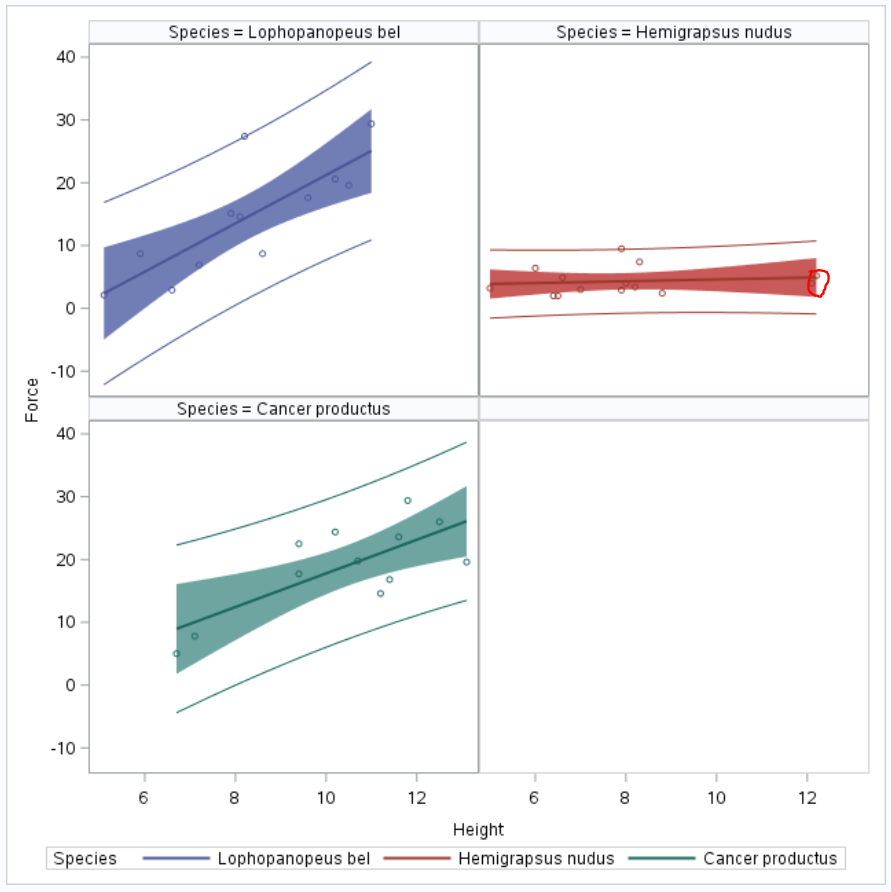
ellipse x=Height y=Force;



**proc** **sgpanel** data=work.crabsraw;

panelby Species;

reg x=Height y=Force / group=Species alpha = .05 CLM CLI;



***Transformed Data (log y)***

/\* transform y (log) \*/

**data** CrabsLogy;

set crabsraw;

LogForce = log(Force);

**proc** **sort** data=work.crabslogy;

by descending Species;

**proc** **sgplot** data=work.crabslogy;

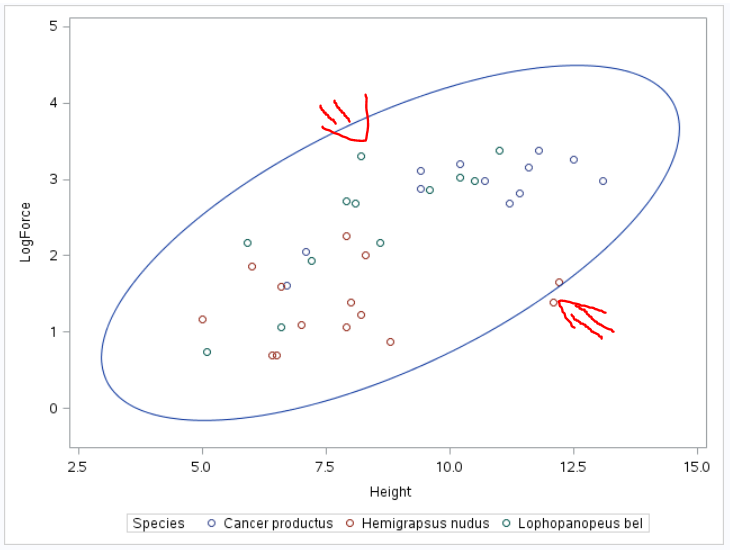
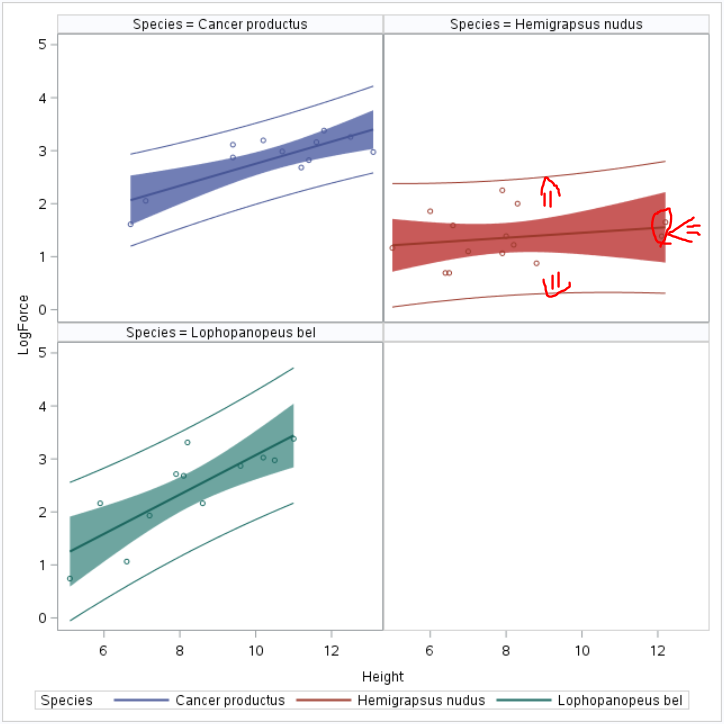
scatter x=Height y=LogForce / group=Species;

ellipse x=Height y=LogForce;

**proc** **sgpanel** data=work.crabslogy;

panelby Species;

reg x=Height y=LogForce / group=Species alpha = **.05** CLM CLI;

1. Step 2: Build a model. (Simply write an appropriate equation as was shown in class.) This model should allow for separate fits (separate lines) for each crab species and should also allow for each line to have its own slope. Use lopho crab as the reference. (This is the default if the data is in alphabetical order.)

***For Species=Lophopanopeus:***

*=*

*=*

***=***

***For Species= Hemigrapsus:***

*=*

***=***

***For Species= Cancer:***

*=*

***=***

1. Step 3: Fit the model. (Fill in the relevant betas in your equation for step 2.) Make sure you provide relevant code and the table of parameter estimates as well.

**proc** **glm** data=work.crabslogy plots = all;

class Species;

Model LogForce = Height Species / Solution clparm alpha = **.05**;

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 0.872473884 | B | 0.38784163 | 2.25 | 0.0311 |
| **Height** | 0.187574342 |  | 0.04353797 | 4.31 | 0.0001 |
| **Species Cancer productus** | 0.014494813 | B | 0.22883811 | 0.06 | 0.9499 |
| **Species Hemigrapsus nudus** | -1.006076015 | B | 0.20107264 | -5.00 | <.0001 |
| **Species Lophopanopeus bel** | 0.000000000 | B | . | . | . |

***For Species=******Lophopanopeus:***

***=***

***For Species=*** ***Hemigrapsus:***

***=***

***For Species= Cancer:***

***=***

1. Step 4: Provide a residual plot, studentized residual plot, histogram of residuals, and q-q plot of residuals to provide evidence of the appropriateness of the model. Provide a short one- or two-sentence discussion of EACH plot.

**For overall data the residuals look good; the spread looks fairly random and there is no clear pattern; though there are outliers (in RStudent Hemigrapssus). I’ll have to leave the observations as is since I cannot determine if they are errors.**

**We have slight curving but overall a straight line in scatter plot of residuals.**

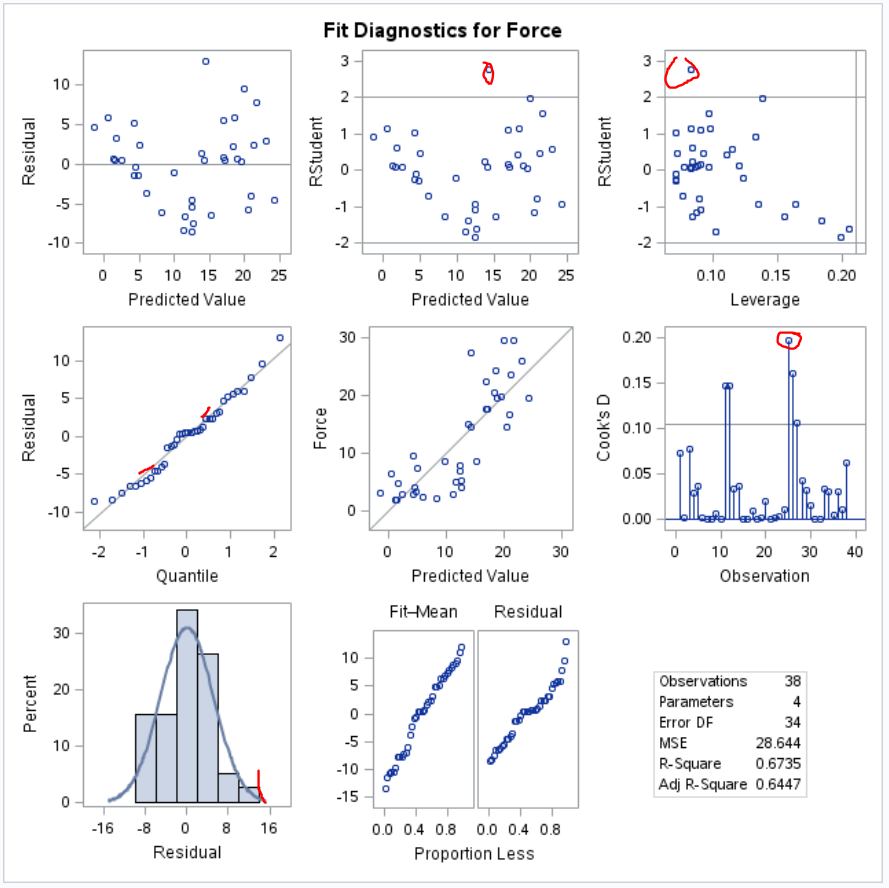
**We have minor right skew in the residuals histogram but it do4esn’t seem too extreme to cause concern.**

**I am assuming that the samples are independent within; Clustering could be an issue since we are dealing with 3 different species (for example each species may react differently to being tested rather than under natural circumstances)**

**proc** **glm** data=work.crabsraw plots = all;

class Species;

Model Force = Height Species / Solution;



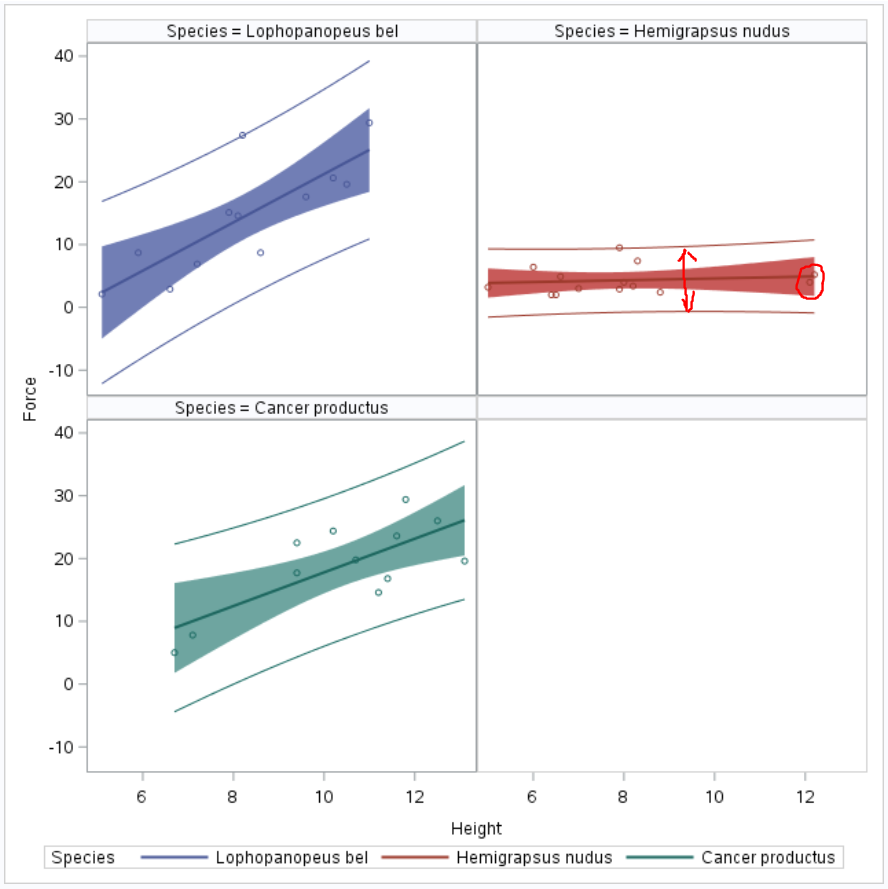
**The slopes are clearly different. Hemigrapsus almost seems to have no correlation between Height and Force based on the plots. This could be caused by the outliers.**

**Also the variance for the Hemigrapsus category seems to be smaller than the others.**

**proc** **sgpanel** data=work.crabsraw;

panelby Species;

reg x=Height y=Force / group=Species alpha = .05 CLM CLI;



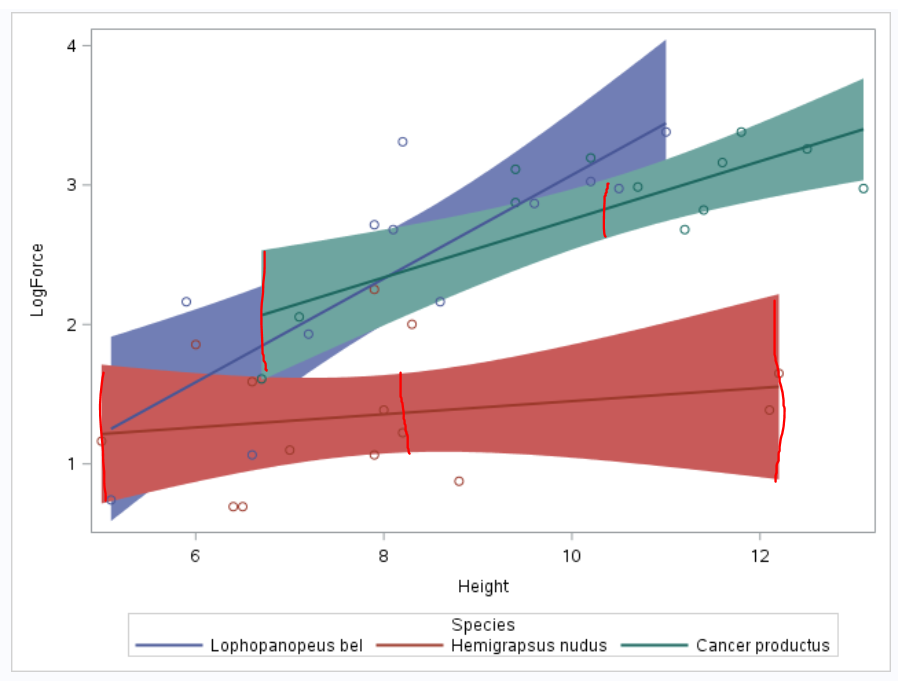
**I found that transforming log(Y) improved this somewhat through trial and error:**

/\* transform y (log) \*/

**data** CrabsLogy;

set crabsraw;

LogForce = log(Force);



1. Step 5: If the fit assessed in Step 4 is sufficient, interpret each coefficient in the model.

**proc** **glm** data=work.crabslogy;

class Species (REF="Lophopanopeus bel");

model LogForce = Species | Height /solution clparm cli alpha = **.05**;

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Intercept** | -0.643462969 | B | 0.61518690 | -1.05 | 0.3034 | -1.896557684 | 0.609631746 |
| **Height** | 0.371510057 | B | 0.07297115 | 5.09 | <.0001 | 0.222872688 | 0.520147427 |
| **Species Cancer productus** | 1.313552082 | B | 0.94795360 | 1.39 | 0.1754 | -0.617366210 | 3.244470374 |
| **Species Hemigrapsus nudus** | 1.622436918 | B | 0.78651500 | 2.06 | 0.0473 | 0.020358297 | 3.224515539 |
| **Species Lophopanopeus bel** | 0.000000000 | B | . | . | . | . | . |
| **Height\*Species Cancer productus** | -0.163131918 | B | 0.09978298 | -1.63 | 0.1119 | -0.366383197 | 0.040119361 |
| **Height\*Species Hemigrapsus nudus** | -0.324387159 | B | 0.09446111 | -3.43 | 0.0017 | -0.516798149 | -0.131976169 |
| **Height\*Species Lophopanopeus bel** | 0.000000000 | B | . | . | . | . | . |

***The difference between claw height of Lophopanopeus and claw height of Cancer does not appear to be significant. (p-value .1754, 95% confidence interval -0.617366210, 3.244470374. I Noted that CI includes 0). Also the difference between claw pressure of Lophopanopeus and claw pressure of Cancer does not appear to be significant (p-value 0.1119, 95% Confidence Interval -0.366383197, 0.040119361)***

***The difference between claw height of Lophopanopeus and claw height of Hemigrapsus does appear to be significant (p-value 0.0473). A 1 unit decrease in claw height is associated with a -0.324387159 multiplicative change in the median of pressure (p-value 0.0017). A 95% confidence interval for this change is -0.516798149, -0.131976169).***

***Because these are different species inference can only be drawn between the 3 in this study and not all crabs.***

1. Provide three individual regression equations (one for each crab species).

***=***

***Lophopanopeus:***

***=***

***Hemigrapsus:***

***=***

***Cancer:***

***=***

1. Read the introduction to the Mammal Brain Weight data that starts on page 239 (Section 9.1.2). Download the Brain data set from 2DS. We would like to see if gestation length and litter size are associated with brain weight after controlling for different body sizes. That is, we already know that brain size is related to body weight; therefore, we don’t want body size to be a confounding variable. We would like to measure the association of the other variables after taking into account the body size.

Answer this question by performing an analysis by following the 5 steps laid out in the problem above. Remember in step 2 to only include the terms that will help you answer this question of interest (QOI).

* Use alpha = 0.05, where necessary.
* Use R and provide relevant code and output.

1. Bonus
2. How many degrees of freedom were used to estimate the error term (MSE) in question 1?
3. What is the estimate of the MSE in question 1?
4. Repeat 1(a) in R.
5. Repeat 1(c) in R.
6. How many degrees of freedom were used to estimate the error term (MSE) in question 2?
7. What is the estimate of the error (MSE) in question 2?