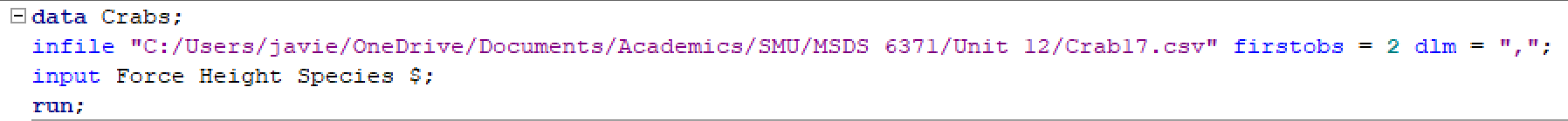
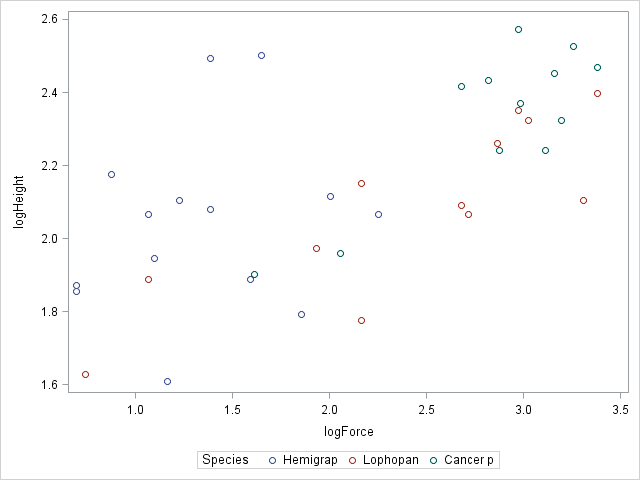
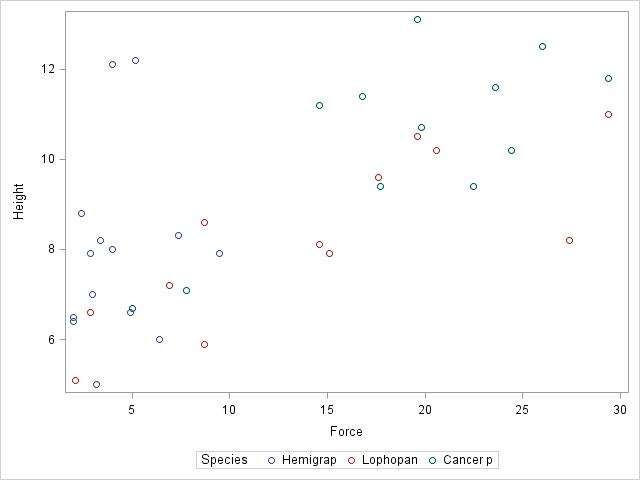
**UNIT 12 HW**

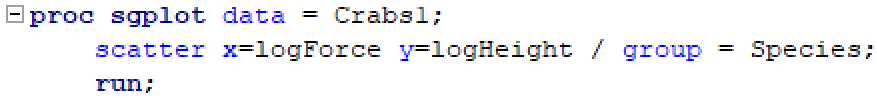
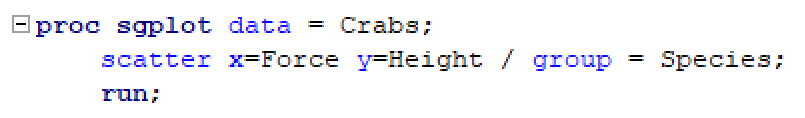
1. As part of a study of the effects of predatory intertidal crab species on snail populations, researchers measured the mean closing forces and the propodus heights of the claws on several crabs of three species. The crab data will be in your files repository.



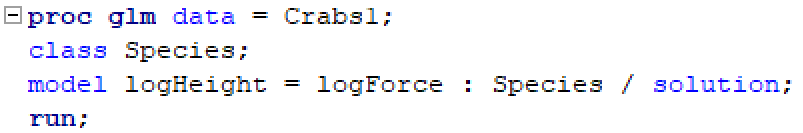
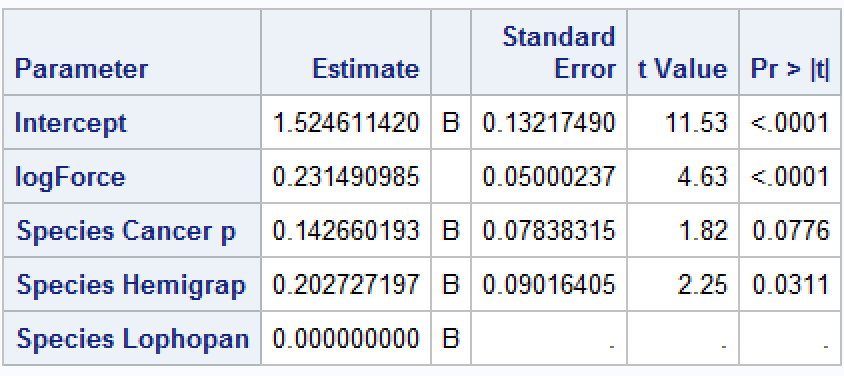
* Use alpha = 0.05, where necessary.
* Use SAS and provide relevant code and output.

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.





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.)
2. 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.



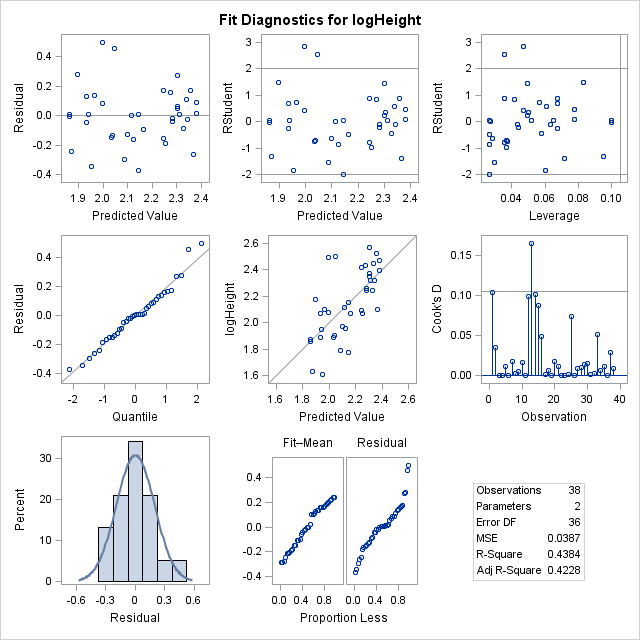
1. Step 4: Provide a residual plot, studentized residual plot, histogram of residuals, and q-q plot of residuals to provide evidence of the appropriateness of the model. Provide a short one- or two-sentence discussion of EACH plot.

Residual Plot: The residuals suggest there is no evidence of a correlation between explanatory variables.

(Studentized) Residual Plot: The residuals are all relatively close to one another and randomly spread out. However, it appears there are 2 potential influential points.

Histogram of Residuals: The histogram provides evidence that the residuals are normally distributed.

Q-Q Plot of Residuals: The influential points are also evident in the QQ plot. We look closer at these points.



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

y-intercept: 1.52 is the mean height assuming force = 0 and species are held constant at Lopho.

log(Force): 100.23 = 1.704 × Force. This means that a 10 fold increase in Force is associated with a 70.4% increase in height, assuming Species are held constant as Lopho.

Species (Cancer): Assuming all other species are held constant, as Force increases, Cancer species increase a 0.14 greater difference than mean Lopho species.

Species (Hemi): Assuming all other species are held constant, as Force increases, Hemi species increase a 0.20 greater than mean Lopho species.

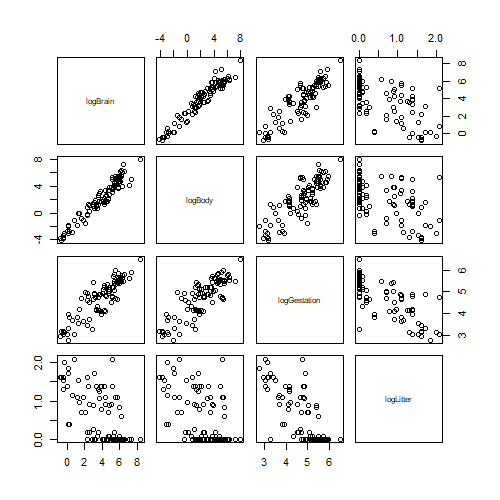
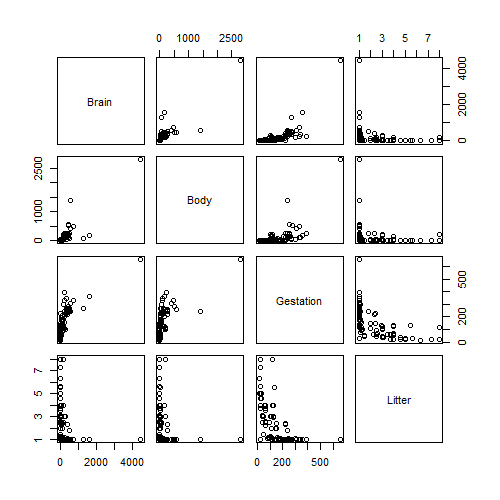
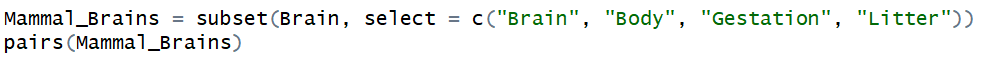
1. Provide three individual regression equations (one for each crab species).
2. Read the introduction to the Mammal Brain Weight data that starts on page 239 (Section 9.1.2). Download the Brain data set from 2DS. We would like to see if gestation length and litter size are associated with brain weight 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.

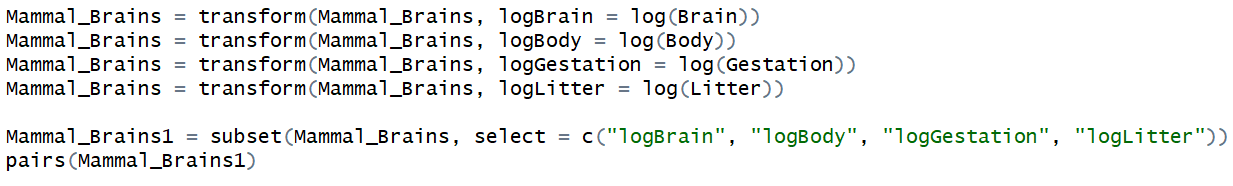


Answer this question by performing an analysis by following the 5 steps laid out in the problem above. Remember in step 2 to only include the terms that will help you answer this question of interest (QOI).

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

Step\_1:



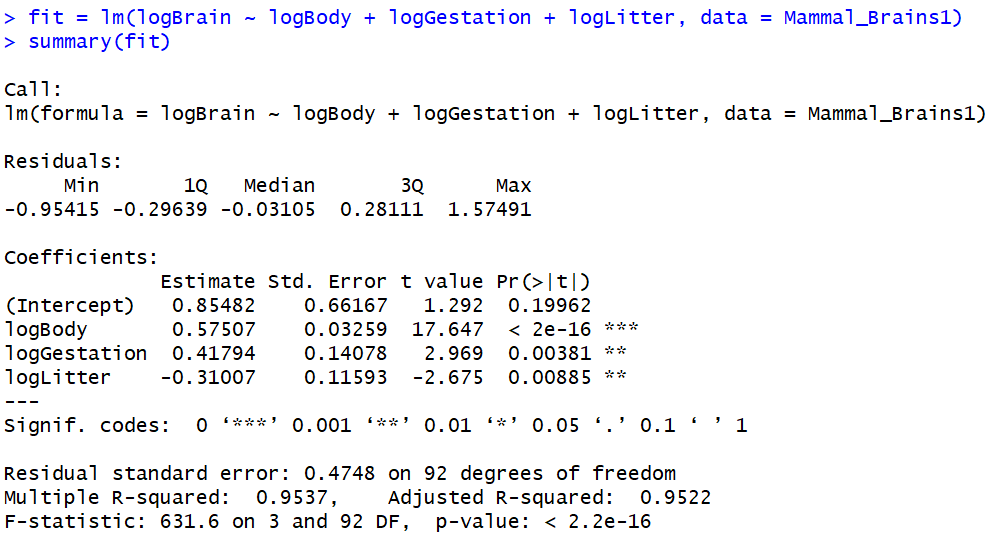


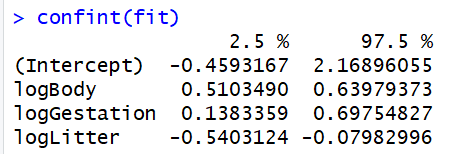
The log/log transformation presents evidence of linearity

Step\_2:

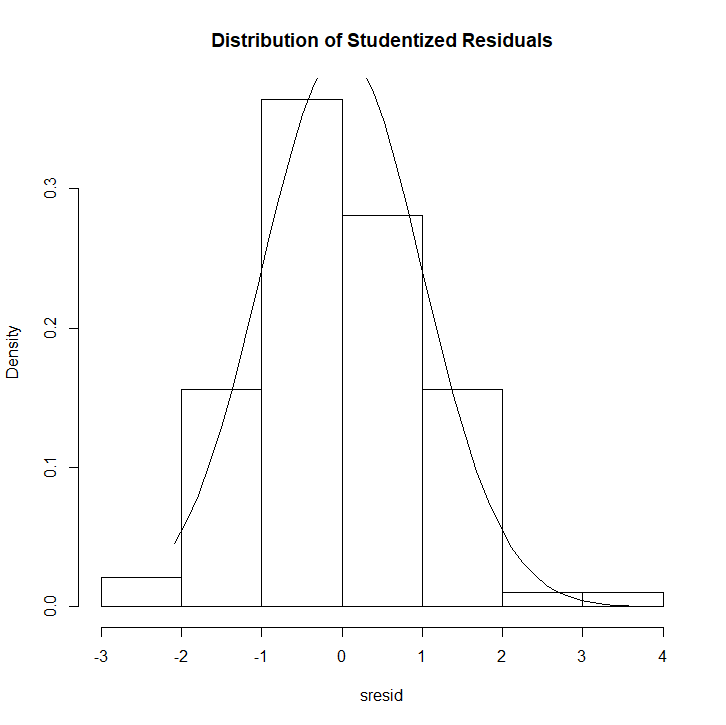
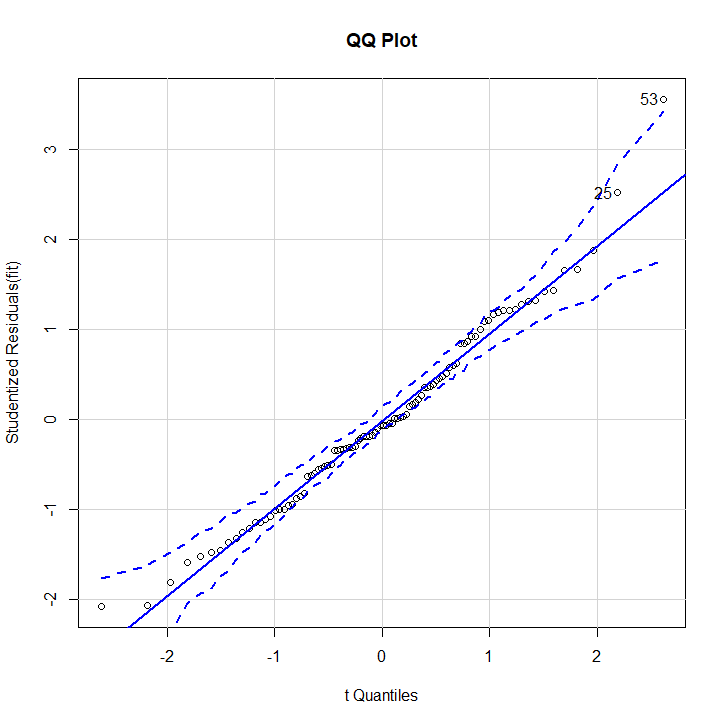


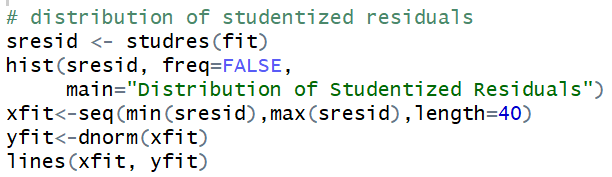
Step\_3:

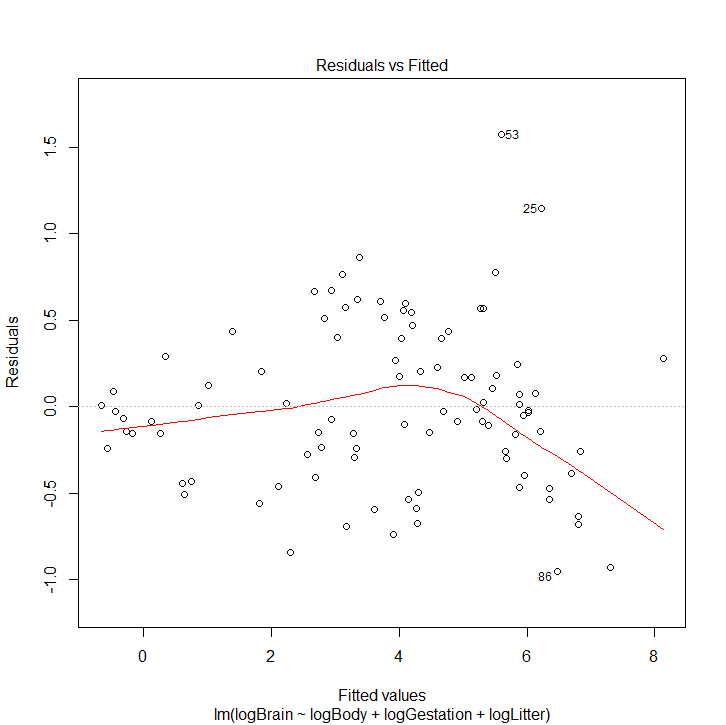
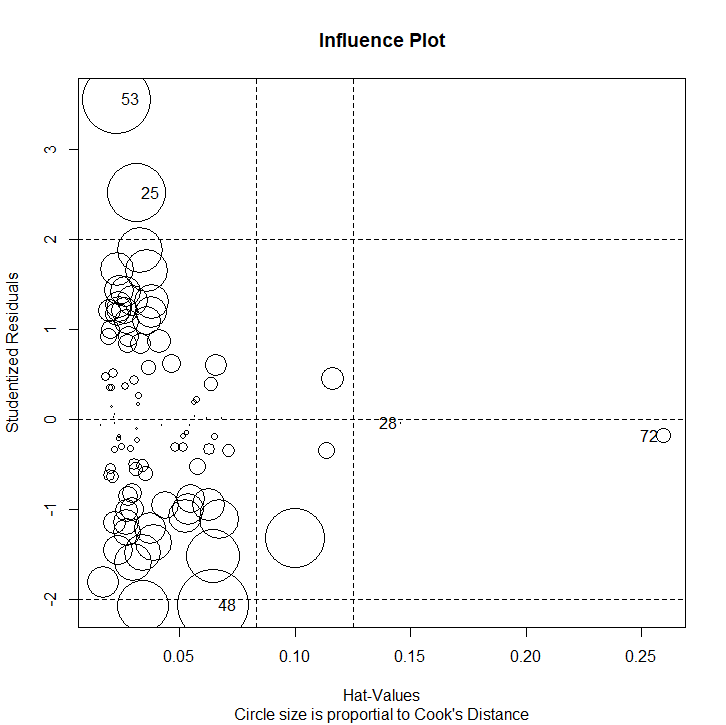




Step\_4:







There is evidence to suggest the residuals are normally distributed based on the residuals histogram. Furthermore, there is concern pertaining to potential outliers and their influence over the mode, as illustrated in the Influence Plot. In addition, the QQ plot for residuals doesn’t present evidence against linearity for the data set. The Residual v. Fitted plot, however suggests there is potential relationship considering the residuals do not appear to be randomized all over the plot.

Step\_5:

There is sufficient evidence to suggest brain weight is associated with gestation and litter size, even after controlling for body size (p-value < 0.001). Based on the data provided, the starting mean brain size for this model is 100.85482 = 7.159g. For every 10 days in gestation period, the median brain size increased by (100.41794 = 2.618) 161%, assuming all else held constant (p-value < 0.00381). However, for every 10 mammals in the litter, the median brain size decreases by 51%, assuming all else is held constant (p-value <0.00885).

1. Bonus
2. How many degrees of freedom were used to estimate the error term (MSE) in question 1?

36

1. What is the estimate of the MSE in question 1?

0.0387

1. Repeat 1(a) in R.
2. Repeat 1(c) in R.
3. How many degrees of freedom were used to estimate the error term (MSE) in question 2?

92

1. What is the estimate of the error (MSE) in question 2?

0.4748