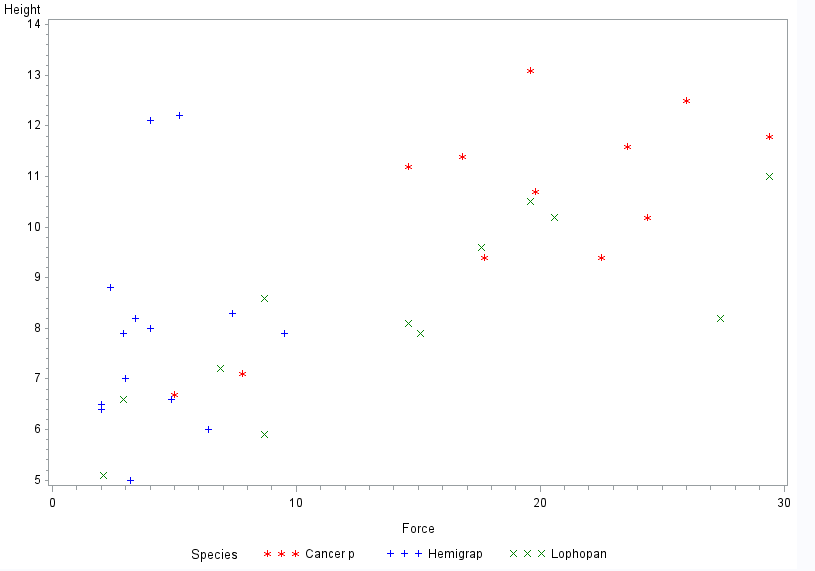
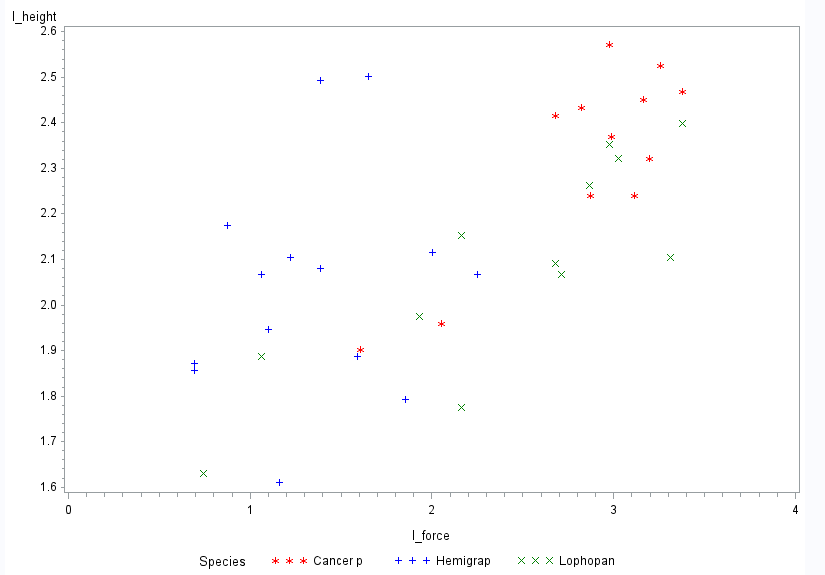
Remy Lagrois

HW 12

Section 402

1. a) Scatterplot for data as is

Lophopan appears to possibly be linear but the other two do not. We will try a log transformation on this data

Data looks better using a log transformation; data points are more evenly distributed. Cancer looks more linear so we will use this for our model

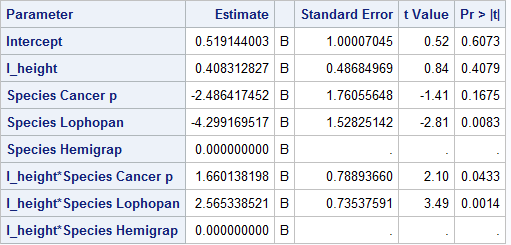
b) Force and height variables are log transformed for this model

µ(force|height, species)

=β0 + β1(height) + β2(cancer) + β3(Loph) + β4(height)(Cancer) + β5(height)(Loph)

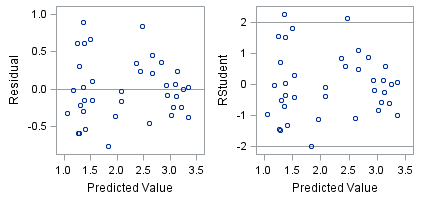
Reference is Hemigrap.

c)

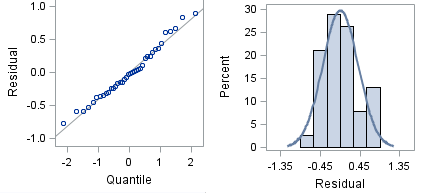


Even though the log(height) term may not be significantly different from zero we should still include it in our model since the interacting terms it is in are significantly different from zero.

d)



The residual plot and studentized residual plot both appear to be mostly random clouds which confirm that using a linear model for the log transformed data was appropriate. The studentized table shows there are two or three outliers which may be influencing our model. However we have no reason to not include these datapoints.



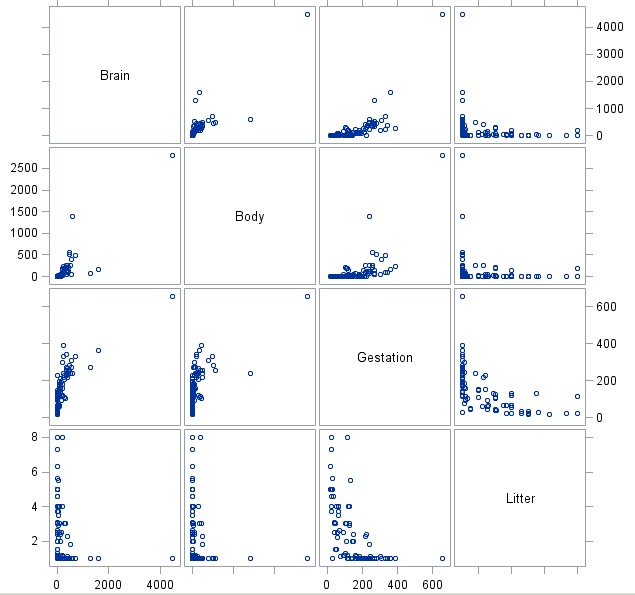
Both the QQ plot and the histogram suggest that the data is normally distributed. The histogram also provides no evidence against our assumption that standard deviates are constant. Overall none of our assumptions seem to be violated.

e) The coefficient for height is not significantly different from zero but we are including it since the interaction terms are significantly different from zero. Given that it is small (0.408) but positive which means that force increases as height increases. The intercept for both Cancer and Lophopan are lower than that for Hemigrap. However the slope for both Cancer and Lophopan are increased in comparison to Hemigrap meaning that height more severely affects their claw force. Between Cancer and Lophopan, Lophopan is more highly affected by height since its interaction coefficient is 2.565 compared to Cancers 1.660.

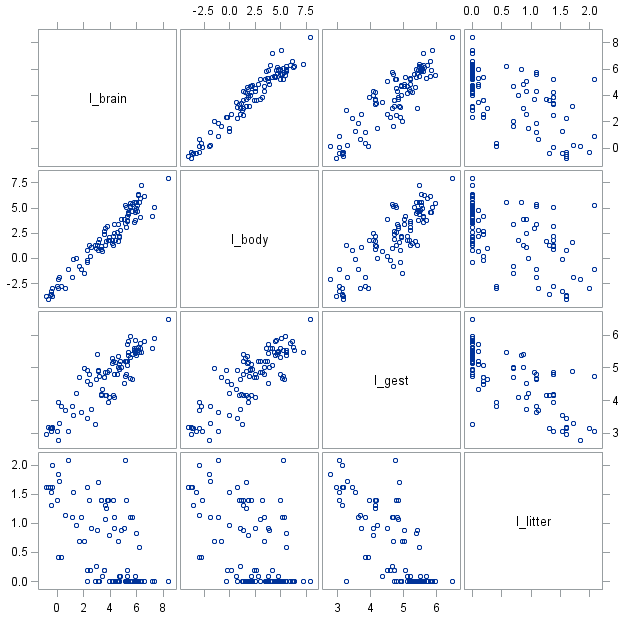
Bonus 1: 32 degrees of freedom where used since there are 5 variables (height, 2 of the species, and the 2 interaction terms)

Bonus 2: The MSE estimate is 0.187

1. a)



None of the variables appear to have linear relationships with brain size (top row of the matrix). We will try a log transformation.

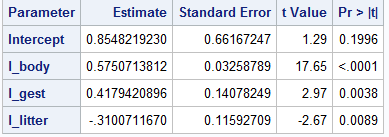


With a log transformation both body and gestation appear to be linearly related to brain size while litter size still does not.

b) µ(brain size|gestation, litter, body) = β0 + β1(gestation) + β2(litter) + β3(body)

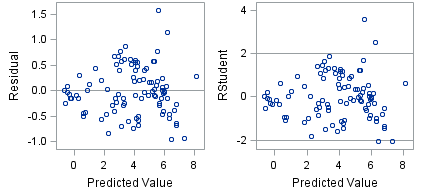
Brain size, litter size, and gestation period are all log transformed in this model.

c)

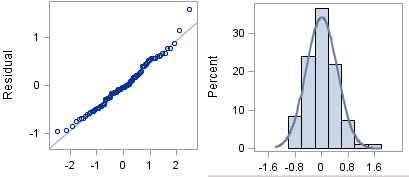


All of our terms are significantly different from zero.

d)



The residual plots are evenly distributed random clouds showing a linear model is a good option. The Studentized plot again shows some outliers though we have no evidence suggesting they should be removed.



The QQ plot and histogram show that our data is normally distributed. There is a slight amount of right skew in the histogram but it is not extreme and doesn’t provide enough evidence that standard deviation is not consistent. Overall we can say that there is not enough evidence to say our assumptions have been violated.

e) Our intercept has a high p-value (.1996) so we fail to reject the null hypothesis that it is different than zero. As such it will not be included in our final model. All our other terms are significantly different from zero and so will be included in the model. The coefficients for body size and gestation are both positive. As gestation length increases and body size increases so does brain size. However litter size has a negative slope which means a larger litter results in a smaller brain size. Our model has an R2 of 0.953 while a model using only body size has an R2 of .92. This means that accounting for litter size and gestation period explain brain size more than body size alone but body size alone is quite influential. This is reflected in the larger slope for body size (0.575) than gestation period (.418).

Bonus 1: The degrees of freedom used was 92 since 3 variables were used (body size, gestation, and litter size).

Bonus 2: The MSE estimate was 0.225