Name: James Tsai
Section: MSDS6371-401
Date: 11/24/15

Question 1.

Step 1: Scatterplot	
Crab Data (no transformation)	**Crab Data (log-transformed)**
Since the log-transformed data is more linear, we proceed with the analysis using log-transformed data.	

Step 2: Build a model
$\mu\{\log(force)\vert height, species\} = \beta_0 + \beta_1 \log(height) + \beta_2 cp + \beta_3 lb + \beta_4 \log(height) * cp + \beta_5 \log(height) * lb$
Note: Hemigrapsus nudus (nb) is not listed since it is the reference variable; when Cancer Productus (cp) and Lophopanopeus Bellus (lb) is set to zero.

Step 3: Fit the model using SAS

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	0.519144003	B	1.00007045	0.52	0.6073
lheight	0.408312827	B	0.48684969	0.84	0.4079
Species Cancer_productus	-2.486417452	B	1.76055648	-1.41	0.1675
Species Lophopanopeus_bellus	-4.299169517	B	1.52825142	-2.81	0.0083
Species Hemigrapsus_nudus	0.000000000	B	.	.	.
lheight*Species Cancer_productus	1.660138198	B	0.78893660	2.10	0.0433
lheight*Species Lophopanopeus_bellus	2.565338521	B	0.73537591	3.49	0.0014
lheight*Species Hemigrapsus_nudus	0.000000000	B	.	.	.

Step 4: Provide a residual plot, studentized residual plot, histogram of residuals and qq plot of residuals

Fit Diagnostics for lforce

Observations	38
Parameters	6
Error DF	32
MSE	0.1874
R-Square	0.7945
Adj R-Square	0.7624

- The scatter plots look random, and thus there is no strong evidence against linearity.
- The scatter plots show several questionable points, but most of the data looks to have equal standard deviation.
- The Q-Q plot shows no strong evidence against normality as the data follows the straight line closely.
- The histogram, while slightly skewed, shows no strong evidence against normality.

Step 5: Interpret each coefficient in the model

- With each increase of multiplicative factor of 2 in height, and the species is Hemigrapsus Nudus, the median force for the crab is associated with a 32 percent increase. ($2^{.408} = 1.32$).
- With each increase of multiplicative factor of 2 in height, and the species is Cancer Productus, the median force for the crab is associated with a 25 percent decrease. ($2^{-.412} = 0.751$).
- With each increase of multiplicative factor of 2 in height, and the species is Lophopanopeus Bellus, the median force for the crab is associated with a 60 percent decrease. ($2^{-1.326} = 0.398$).

BONUS: How many degrees of freedom were used to estimate the error term (MSE)?

The degrees of freedom used to estimate the MSE is 32.
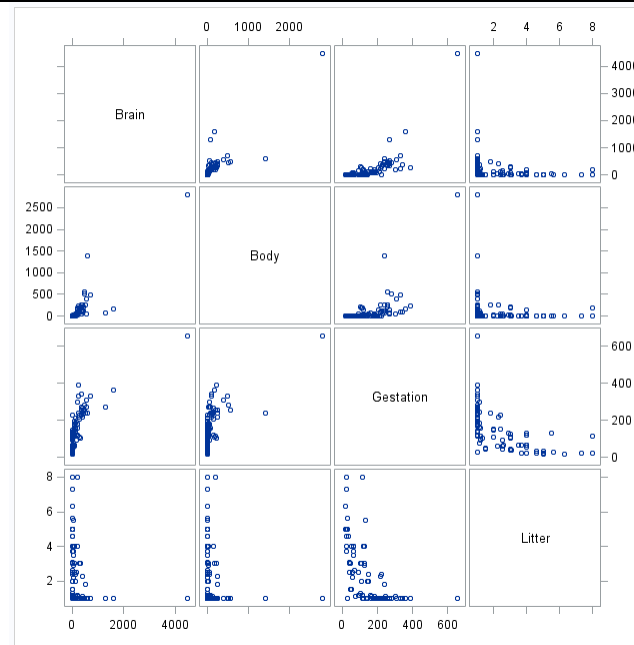
BONUS: What is the estimate of the MSE?

The estimate of the MSE is 0.18741.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23.19217	4.63843	24.75	<.0001
Error	32	5.99713	0.18741		
Corrected Total	37	29.18930			

Question 2.

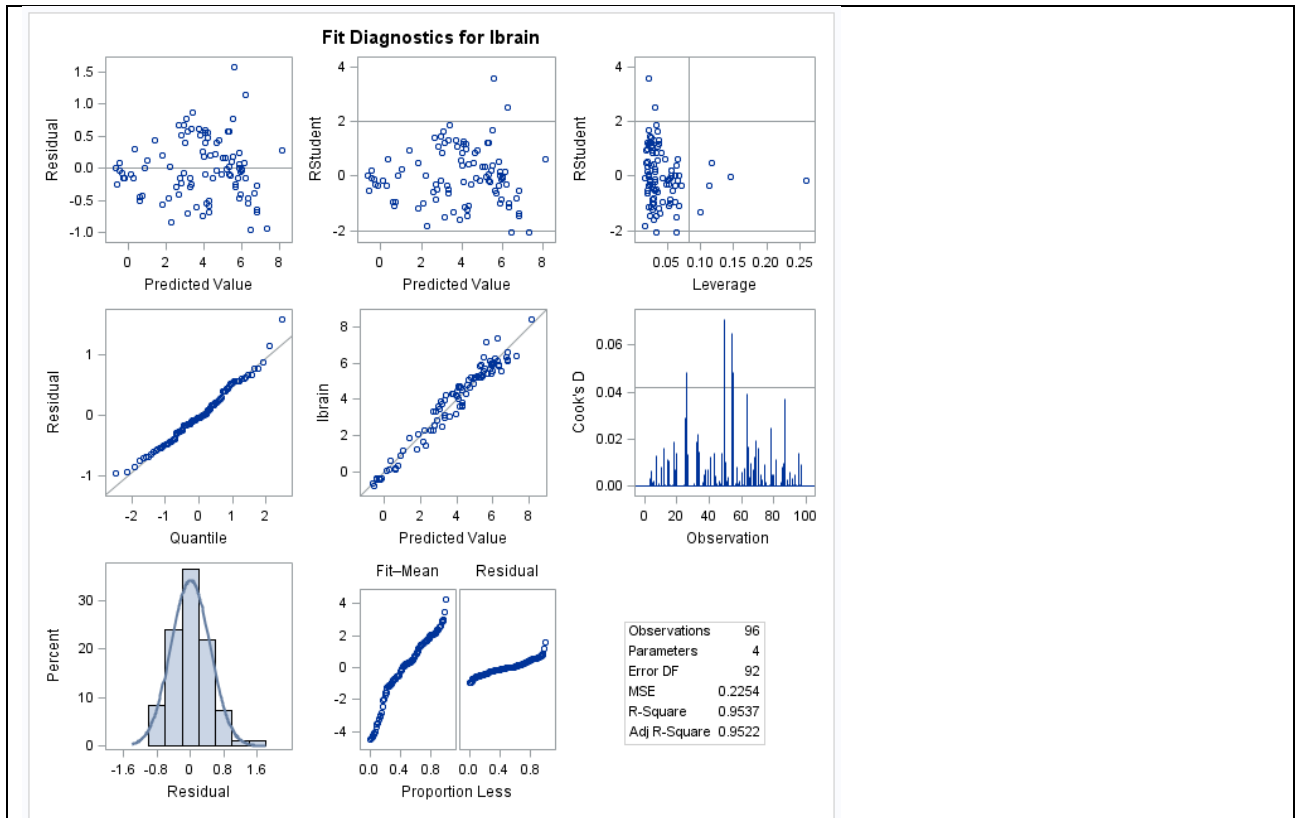| Step 1: Scatterplot	
Brain Data (no transformation)	**Brain Data (log-transformed)**
The original matrix scatter plot is extremely non-linear. We proceed with the log-transformed data, as it is much closer to linearity.	
Step 2: Build a model	
$\mu\{\log(brain)|body, gestation, litter\} = \beta_0 + \beta_1 \log(body) + \beta_2 \log(gestation) + \beta_3 \log(litter)$	
Step 3: Fit the model using SAS	
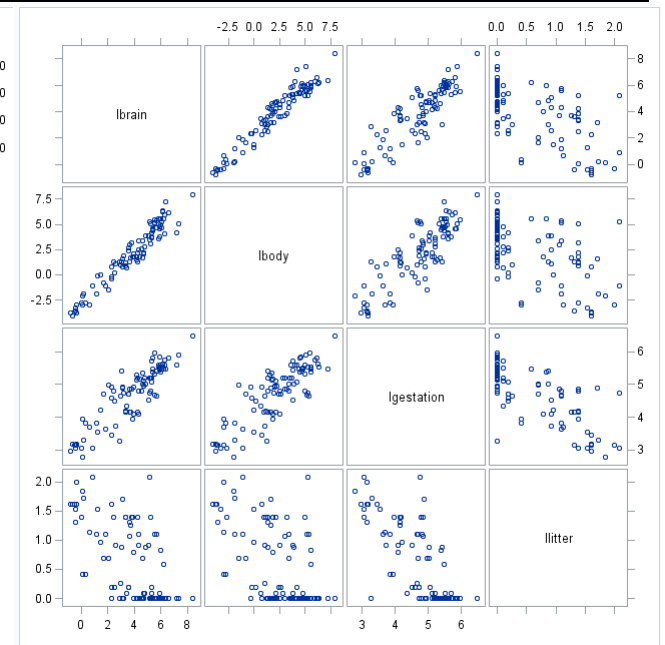
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.8548219230	0.66167247	1.29	0.1996
lbody	0.5750713812	0.03258789	17.65	<.0001
lgestation	0.4179420896	0.14078249	2.97	0.0038
llitter	-.3100711670	0.11592709	-2.67	0.0089

Step 4: Provide a residual plot, studentized residual plot, histogram of residuals and qq plot of residuals

Fit Diagnostics for lbrain

Observations	96
Parameters	4
Error DF	92
MSE	0.2254
R-Square	0.9537
Adj R-Square	0.9522

- The scatter plots look random, and thus there is no strong evidence against linearity.
- The scatter plots show several questionable points, but most of the data looks to have equal standard deviation.
- The Q-Q plot shows no strong evidence against normality as the data follows the straight line closely.
- The histogram shows no strong evidence against normality.

Step 5: Interpret each coefficient in the model
- There is evidence to suggest at $\alpha = 0.05$ that litter size was associated with brain weight after accounting for body weight and gestation (p-value = 0.0089). An increase of a multiplicative factor of 2 in litter size would suggest an estimated multiplicative change of 19% ($2^{-.31}$) decrease in median brain size.   - There is evidence to suggest at $\alpha = 0.05$ that gestation length was associated with brain weight after accounting for body weight and litter size (p-value = 0.0038). An increase of a multiplicative factor of 2 in gestation length would suggest an estimated multiplicative change of 33.6% ($2^{.418}$) increase in median brain size.
BONUS: How many degrees of freedom were used to estimate the error term (MSE)?
The degrees of freedom used to estimate the MSE is 92.
BONUS: What is the estimate of the MSE?
The estimate of the MSE is 0.22539.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	427.07552	142.35851	631.60	<.0001
Error	92	20.73608	0.22539		
Corrected Total	95	447.81160			