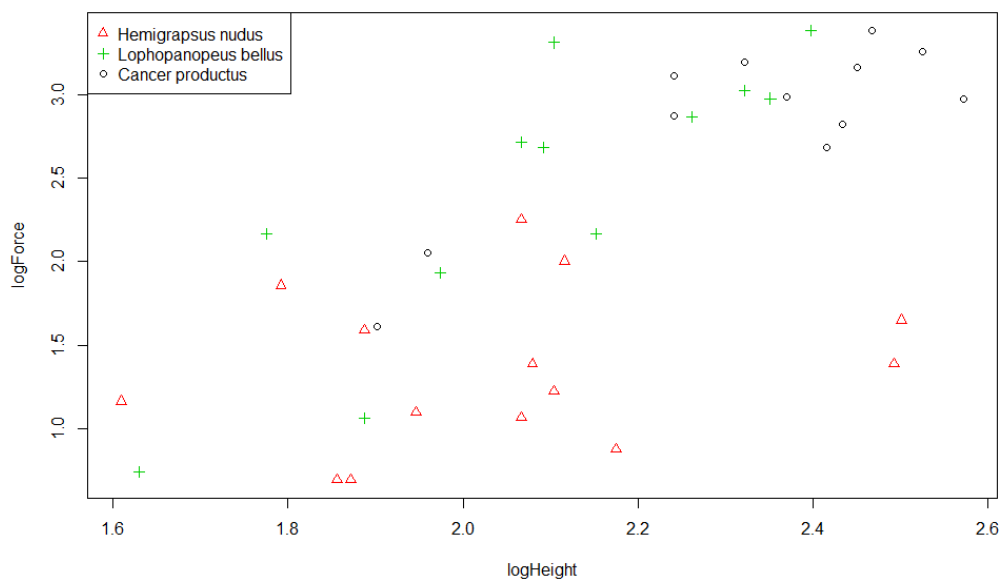
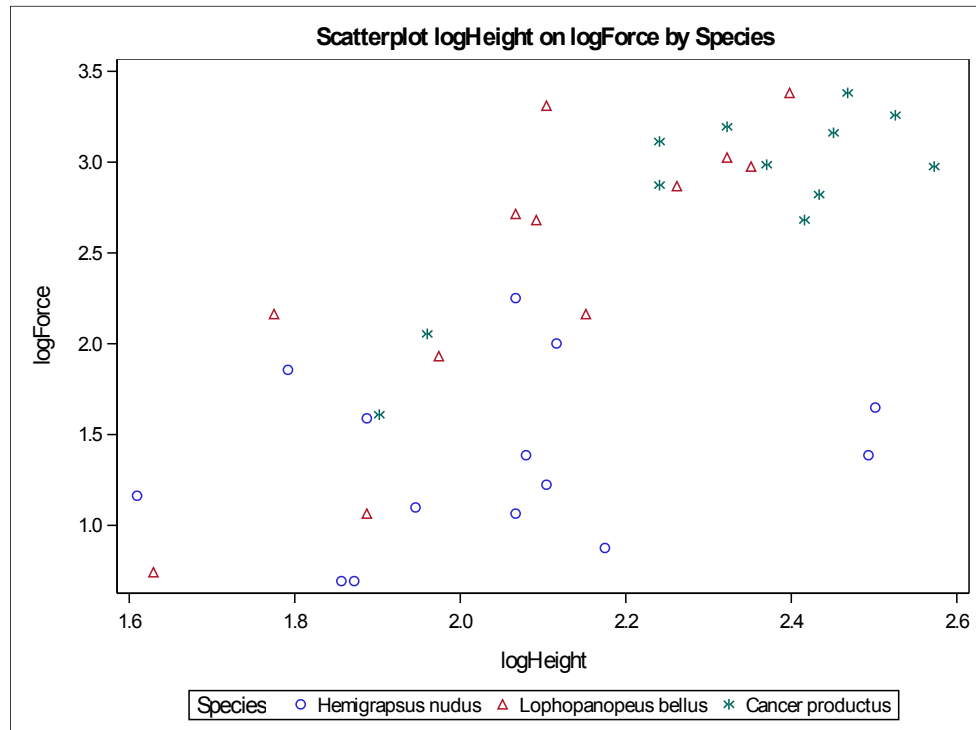
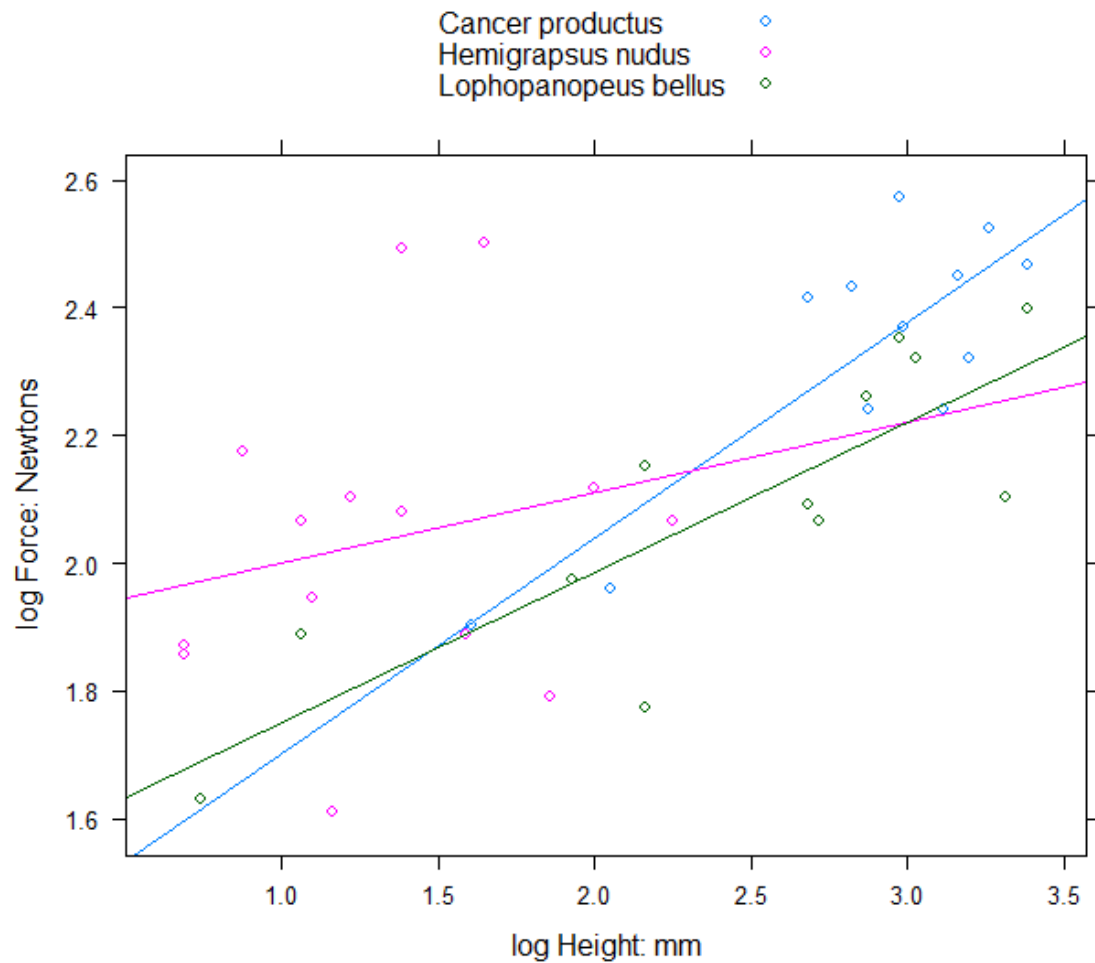


Question 1)**Part A) Scatterplot Assessment – Log-Force vs. Log-Height (SAS/R)**

After assessing various transformations, $\log(\text{Force})$ and $\log(\text{Height})$ were used to make regression estimations of median closing force by propodus height across several crabs of three species.



EDA also produced the following individual plots of logForce on logHeight by Species:



Part B) Model – Log-Force vs. Log-Height, Species *with Interaction* (REF = Lophopanopeus B.)

$$\log\text{Force} = \beta_0 + \beta_1 \log\text{Height} + \beta_2 \text{Species}_C + \beta_3 \text{Species}_H + \beta_4 \text{Species}_L +$$

$$\beta_5 \log\text{Height} * \text{Species}_C + \beta_6 \log\text{Height} * \text{Species}_H + \beta_7 \log\text{Height} * \text{Species}_L + \varepsilon$$

Part C) Model Fit and Parameter Estimate Tables (SAS /R)

SAS

```
proc glm data = crabz plots = diagnostics(label);
class Species;
model logForce = logHeight | Species / solution;
title 'Regression of logForce on logHeight with Interaction: Ref = Lophopanopeus bellus';
run; quit;
```

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-3.780025514	B	1.15560007	-3.27	0.0026
logHeight	2.973651348	B	0.55113983	5.40	<.0001
Species Cancer productus	1.812752064	B	1.85332936	0.98	0.3354
Species Hemigrapsus nudus	4.299169517	B	1.52825142	2.81	0.0083
Species Lophopanopeus bellus	0.000000000	B	.	.	.
logHeight*Species Cancer productus	-0.905200323	B	0.83015267	-1.09	0.2837
logHeight*Species Hemigrapsus nudus	-2.565338521	B	0.73537591	-3.49	0.0014
logHeight*Species Lophopanopeus bellus	0.000000000	B	.	.	.

R

```

> require(Sleuth3)
> require(mosaic)
> options(digits = 4)
> attach(ex0722)
> ex0722$logHeight <- with(ex0722, log(Height))
> ex0722$logForce <- with(ex0722, log(Force))
> logfit_Default <- lm(logForce ~ logHeight*Species)
> newerType <- relevel(Species, ref = "Lophopanopeus bellus")
> logfit_L <- lm(logForce ~ logHeight*newerType)
> summary(logfit_L)

```

Call:

lm(formula = logForce ~ logHeight * newerType)

Residuals:

```

      Min       1Q   Median       3Q      Max
-0.7668 -0.2851 -0.0231  0.2425  0.8882

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -3.780      1.156   -3.27  0.0026 **
logHeight       2.974      0.551    5.40 6.3e-06 ***
newerTypeCancer productus  1.813      1.853    0.98  0.3354
newerTypeHemigrapsus nudus  4.299      1.528    2.81  0.0083 **
logHeight:newerTypeCancer productus -0.905      0.830   -1.09  0.2837
logHeight:newerTypeHemigrapsus nudus -2.565      0.735   -3.49  0.0014 **
---

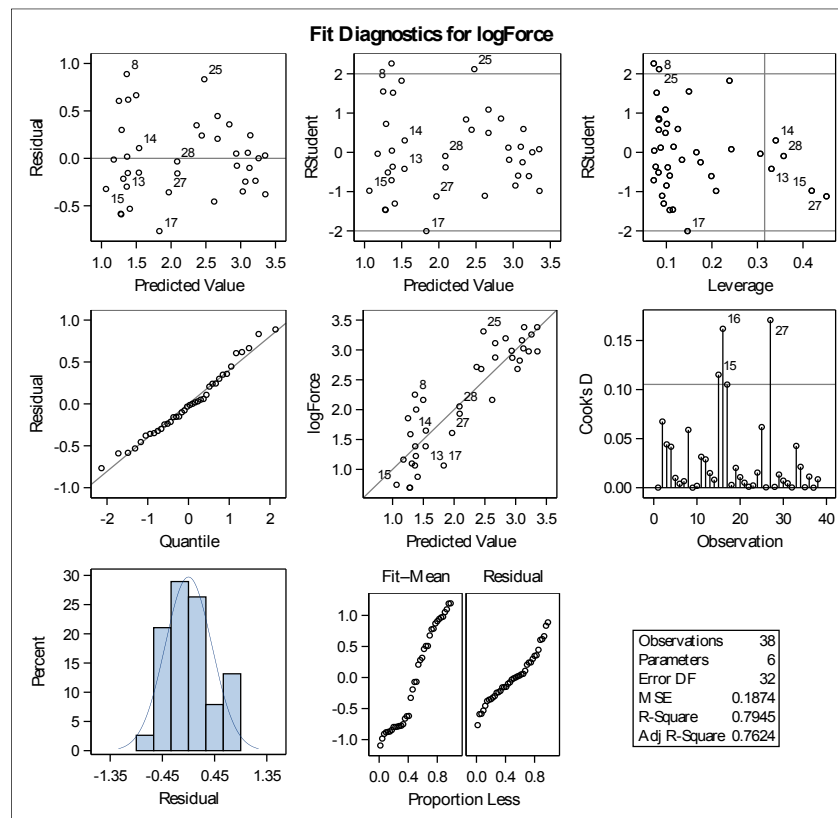
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

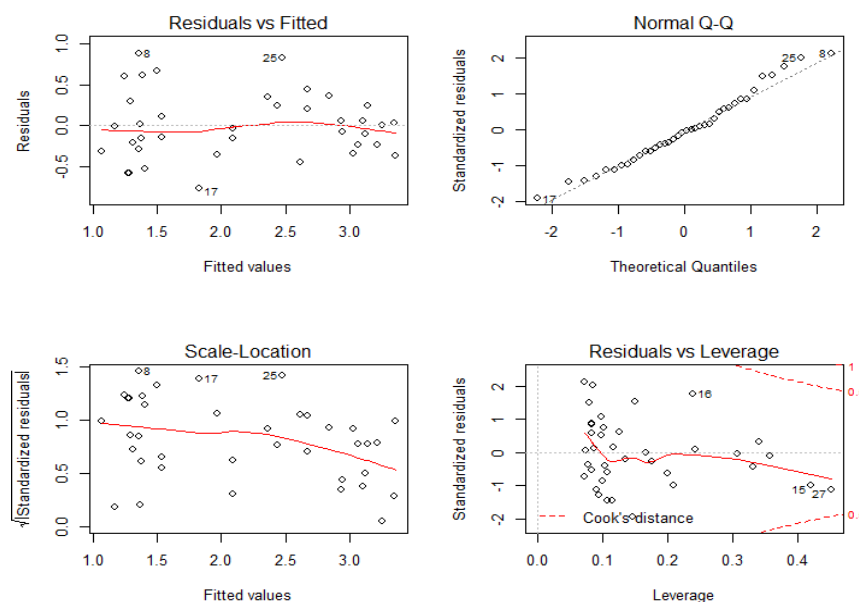
```

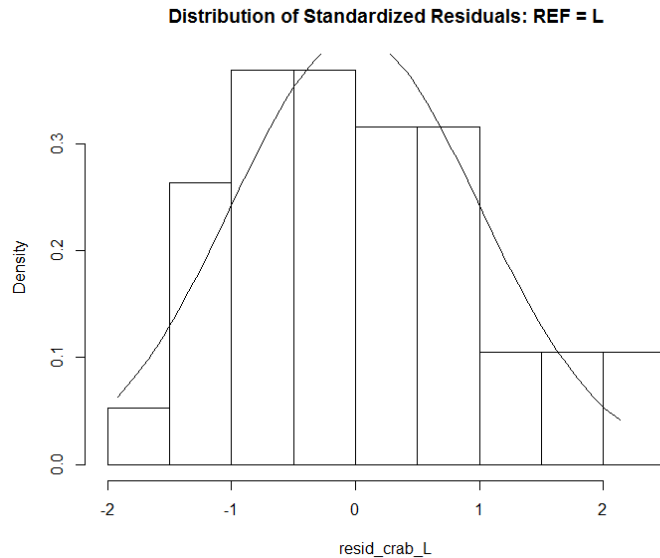
Residual standard error: 0.433 on 32 degrees of freedom
Multiple R-squared:  0.795,    Adjusted R-squared:  0.762
F-statistic: 24.8 on 5 and 32 DF,  p-value: 3.94e-10

```

Part D) Residual Plots and Model Evaluation (SAS /R)

To assess *normality* and *constant variance* in both SAS/R, the **Residual-Predicted Value** plot adequately demonstrates random scatter about the zero line. Furthermore, observations on the **Studentized Residual-Predicted Value** plots remain largely within 2 standard deviations. **QQ-plots** exhibit adequate residual normality. Similarly, **Residual Histograms** exhibit normality under the density curve.





Part E) Model Interpretation

To interpret the parameter coefficients, estimates were identified for three models, setting the reference level to each of three species types.

For REF = Species *Lophopanopeus B.*:

$$\log \text{Force} = -3.78 + 2.9737 \log \text{Height} + 1.8128 \text{Species}_C + 4.2992 \text{Species}_H + (0) \text{Species}_L -$$

$$.9052 \log \text{Height} * \text{Species}_C - 2.5653 \log \text{Height} * \text{Species}_H + (0) \log \text{Height} * \text{Species}_L + \varepsilon$$

$$= -3.78 + 2.9737 \log \text{Height} + 1.8128(0) + 4.2992(0) + (0)(\text{Species}_L) -$$

$$.9052 \log \text{Height} * (0) - 2.5653 \log \text{Height} * (0) + (0) \log \text{Height} * (1) + \varepsilon$$

$$= -3.78 + 2.9737 \log \text{Height} + \varepsilon$$

$$\text{Force} = e^{(-3.78)} * e^{(2.9737 \log \text{Height})} * \varepsilon$$

$$= (0.02282) \text{Height}^{2.9737} * \varepsilon$$

→ The predicted median Force of Species *Lophopanopeus B.* with a Height of 1mm is .02282 N. Doubling the Height multiplies the predicted median Force by ($2^{2.9737}$), or 7.8555, increasing it by 686%.

For REF = Species *Hemigraspus N.*:

$$\log \text{Force} = .5191 + .4083 \log \text{Height} - 2.4864 \text{Species}_C - 4.2992 \text{Species}_L + (0) \text{Species}_H +$$

$$1.6601 \log \text{Height} * \text{Species}_C + 2.5653 \log \text{Height} * \text{Species}_L + (0) \log \text{Height} * \text{Species}_H + \varepsilon$$

$$\begin{aligned}
 &= .5191 + .4083 \log \text{Height} - 2.4864(0) - 4.2992(0) + (0)(\text{Species_H}) + \\
 &1.6601 \log \text{Height} * (0) + 2.5653 \log \text{Height} * (0) + (0) \log \text{Height} * (1) + \varepsilon \\
 &= .5191 + .4083 \log \text{Height} + \varepsilon
 \end{aligned}$$

$$\begin{aligned}
 \text{Force} &= e^{(.5191)} * e^{(.4083 \log \text{Height})} * \varepsilon \\
 &= (1.6805) \text{Height}^{.4083} * \varepsilon
 \end{aligned}$$

→ The predicted median Force of Species *Hemigraspus N.* with a Height of 1mm is 1.6805 N. Doubling the Height multiplies the predicted median Force by ($2^{.4083}$), or 1.3271, increasing it by 33%.

For REF = Species Cancer P.:

$$\begin{aligned}
 \log \text{Force} &= -1.9673 + 2.0685 \log \text{Height} + 2.4864 \text{Species_H} - 1.8128 \text{Species_L} + (0) \text{Species_C} - \\
 &1.6601 \log \text{Height} * \text{Species_H} + .9052 \log \text{Height} * \text{Species_L} + (0) \log \text{Height} * \text{Species_C} + \varepsilon \\
 &= -1.9673 + 2.0685 \log \text{Height} - 2.4864(0) - 1.8128(0) + (0)(\text{Species_C}) - \\
 &1.6601 \log \text{Height} * (0) + .9052 \log \text{Height} * (0) + (0) \log \text{Height} * (1) + \varepsilon \\
 &= -1.9673 + 2.0685 \log \text{Height} + \varepsilon
 \end{aligned}$$

$$\begin{aligned}
 \text{Force} &= e^{(-1.9673)} * e^{(2.0685 \log \text{Height})} * \varepsilon \\
 &= (0.1398) \text{Height}^{2.0685} * \varepsilon
 \end{aligned}$$

→ The predicted median Force of Species *Cancer P.* with a Height of 1mm is .1398 N. Doubling the Height multiplies the predicted median Force by ($2^{2.0685}$), or 4.1945, increasing it by 319%.

Bonus Question – Part A)

From the ANOVA tables in SAS/R, the **degrees of freedom** used to estimate the error term (MSE) were **df = 32**.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23.19216955	4.63843391	24.75	<.0001
Error	32	5.99713447	0.18741045		
Corrected Total	37	29.18930402			

R-Square	Coeff Var	Root MSE	logForce Mean
0.794543	20.04661	0.432909	2.159514

```
> anova(logfit_L)
```

Analysis of Variance Table

Response: logForce

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
logHeight	1	12.80	12.80	68.29	1.9e-09	***
Species	2	8.01	4.00	21.37	1.3e-06	***
logHeight:Species	2	2.38	1.19	6.36	0.0047	**
Residuals	32	6.00	0.19			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> summary(logfit_L)
```

...

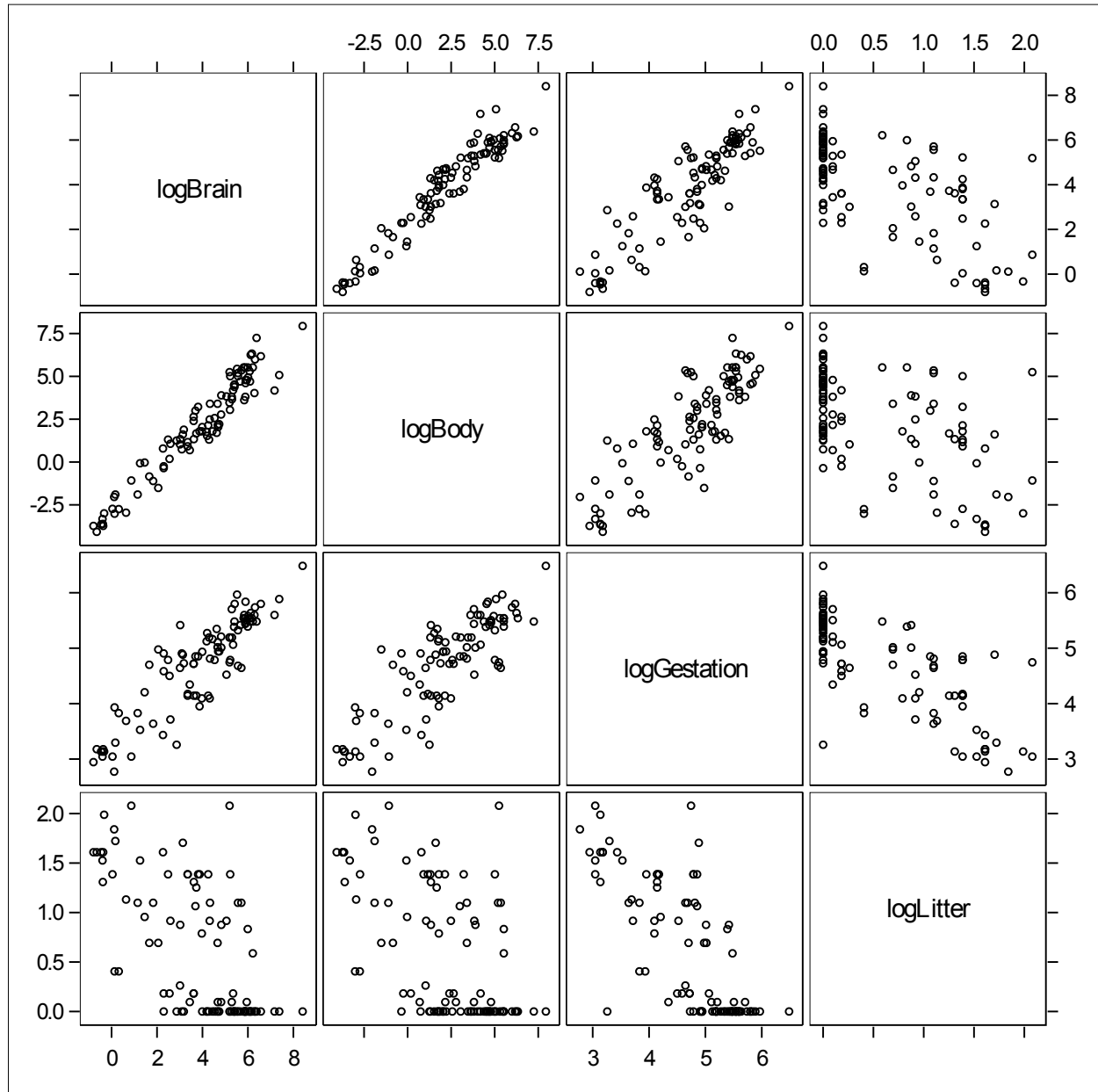
Residual standard error: 0.433 on 32 degrees of freedom
Multiple R-squared: 0.795, Adjusted R-squared: 0.762
F-statistic: 24.8 on 5 and 32 DF, p-value: 3.94e-10

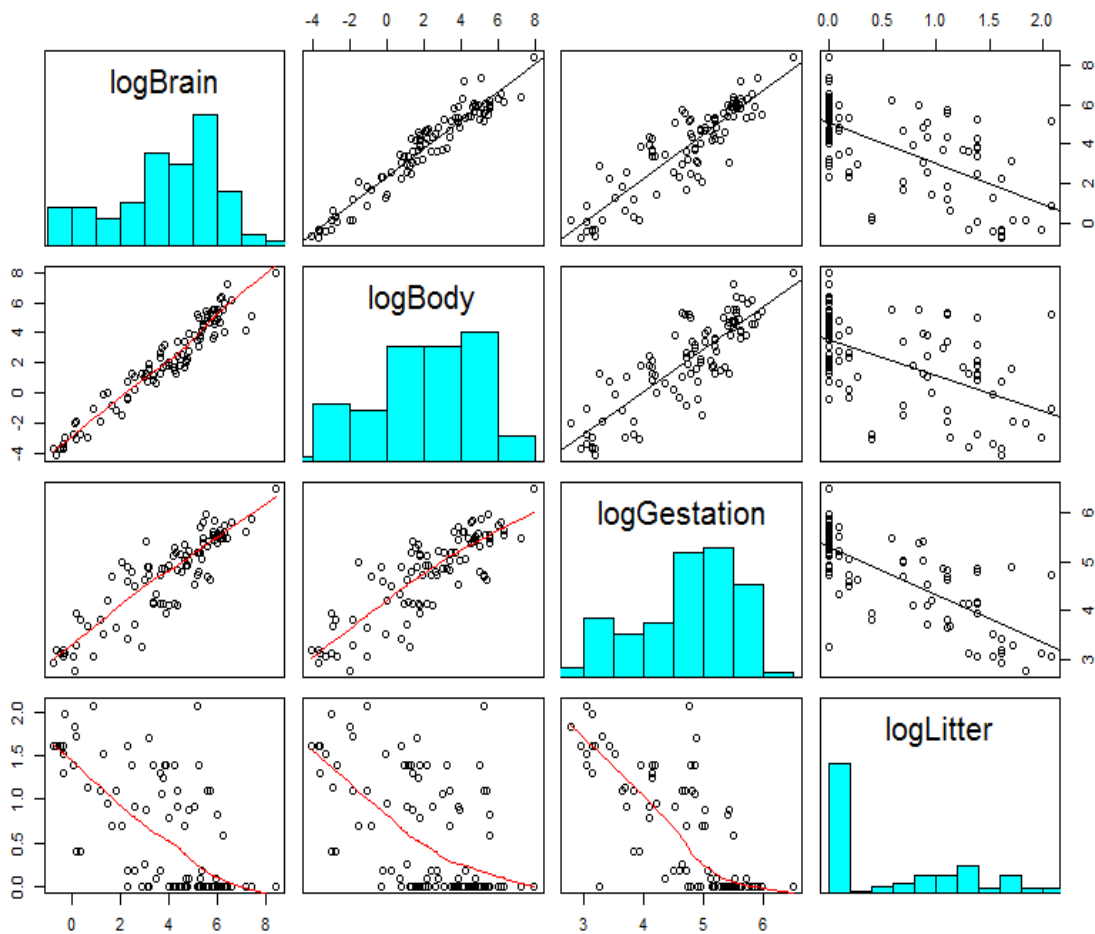
Bonus Question – Part B)

Similarly, from the ANOVA tables in SAS/R, the **estimate of the MSE** was .18741; the Root MSE (residual standard error) was 0.4329.

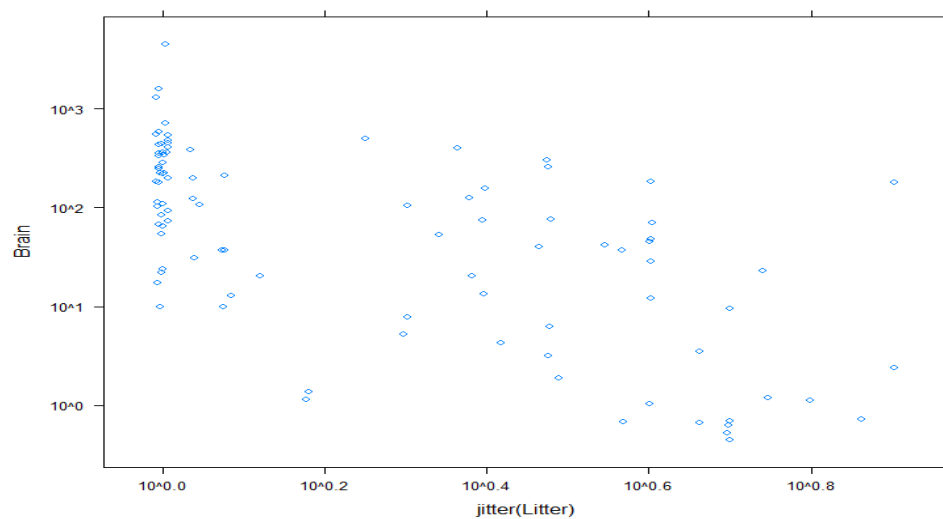
Question 2)**Part A) Scatterplot Assessment – Log-Brain vs. Log-Body, Log-Gestation, and Log-Litter (SAS/R)**

After assessing various transformations, $\log(\text{Brain})$ and $\log(\text{Body})$, $\log(\text{Gestation})$, and $\log(\text{Litter})$ were used to make regression estimations of median Brain weight by Body size, Gestation Period, and Litter size for 96 mammalian species.

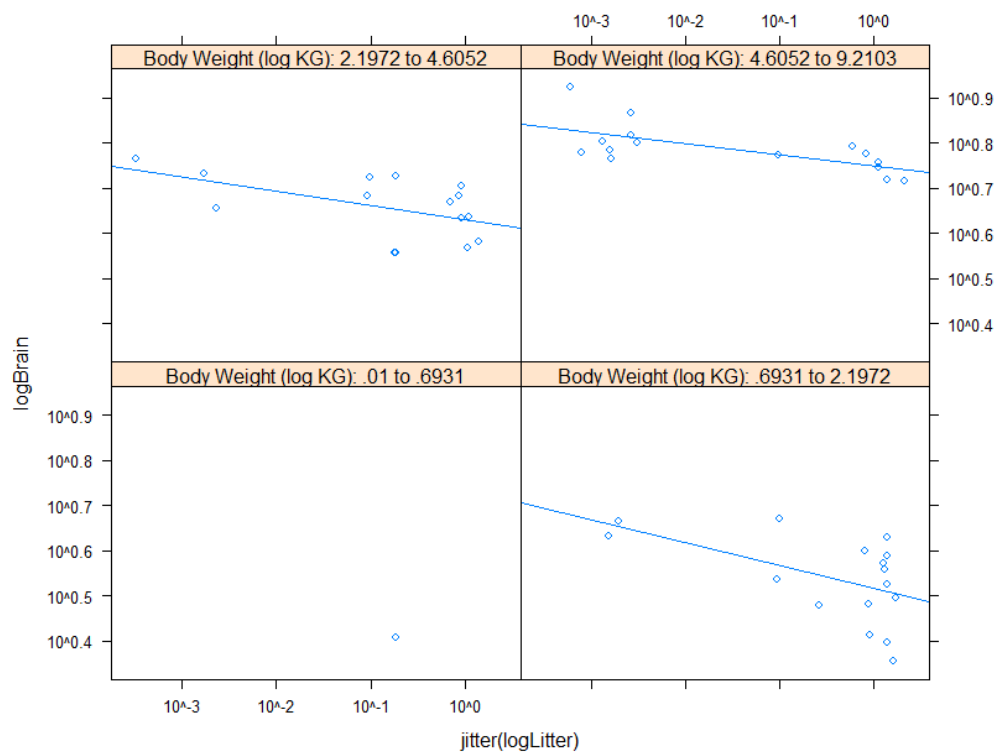
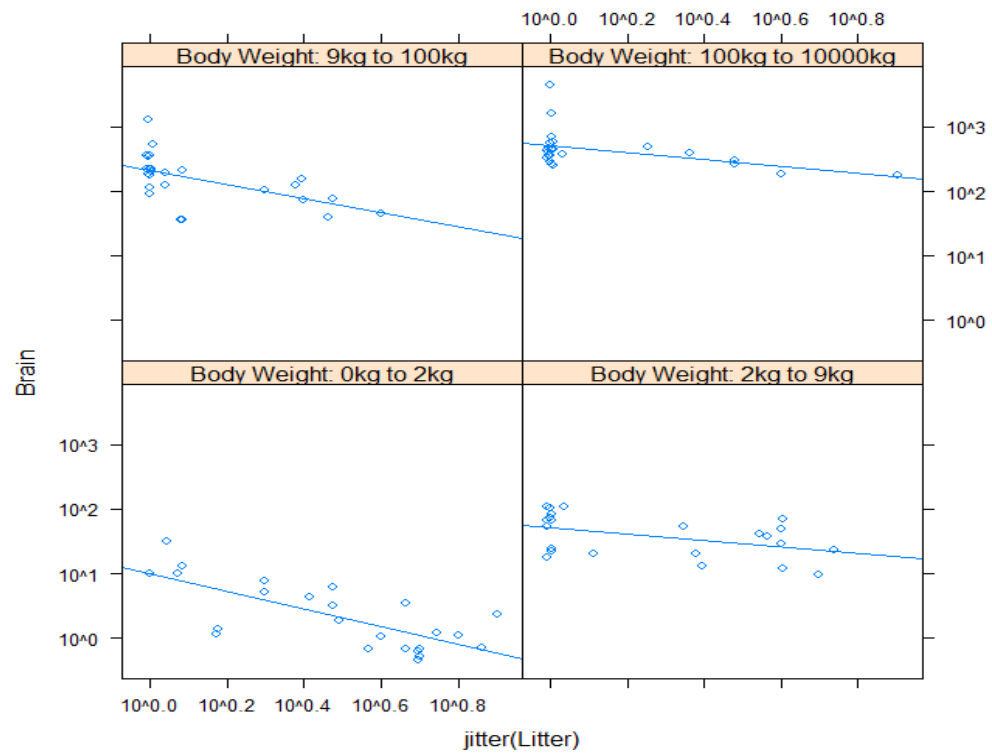




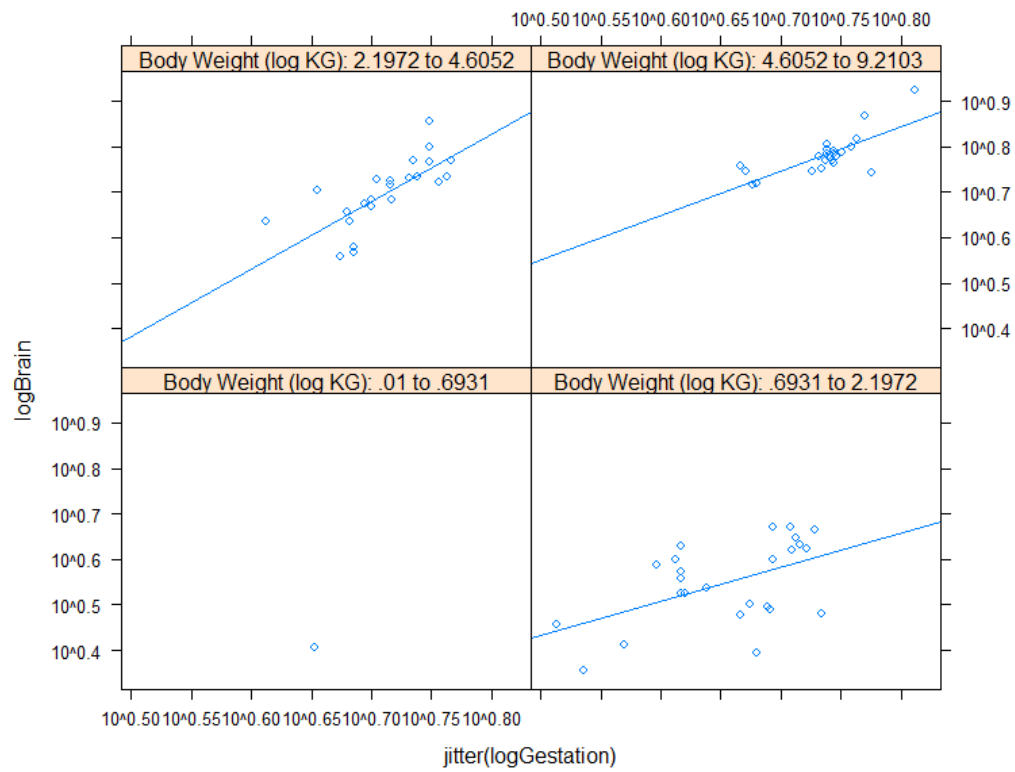
To explore the linear contribution of logLitter, and to evaluate the known relation between Brain size and Body weight, EDA produced the following jitter plots, trellis graphs, and component residual plots, as well as individual plots of logBrain on logGestation and logLitter by Body weight group :



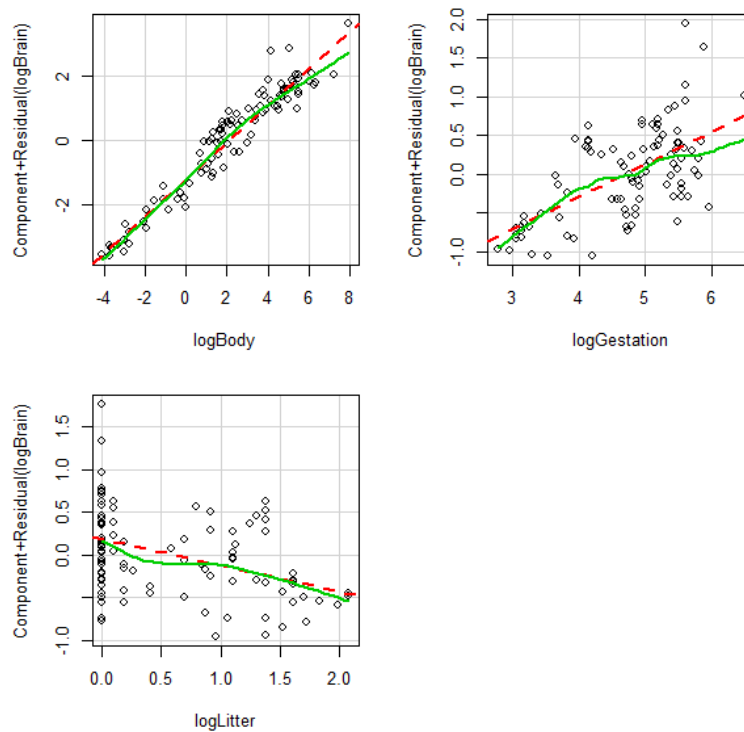
4/9/17

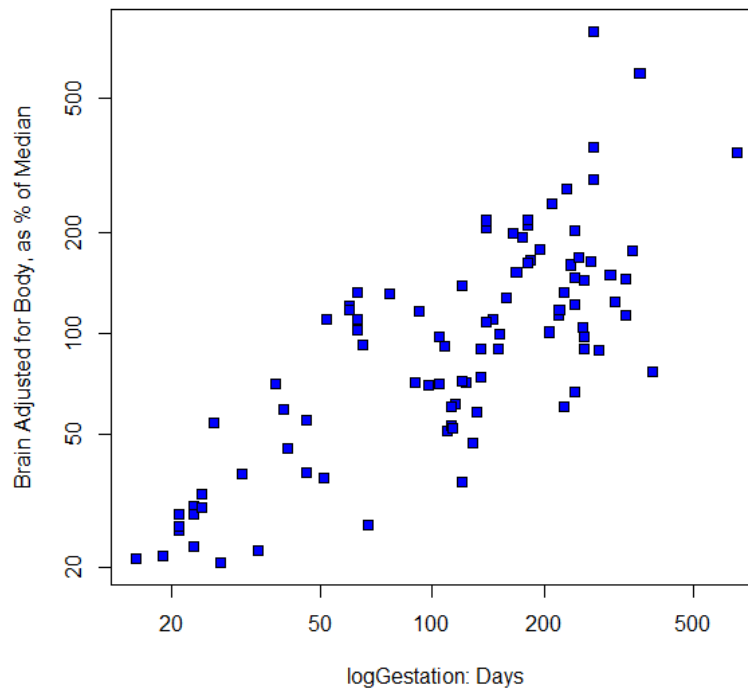
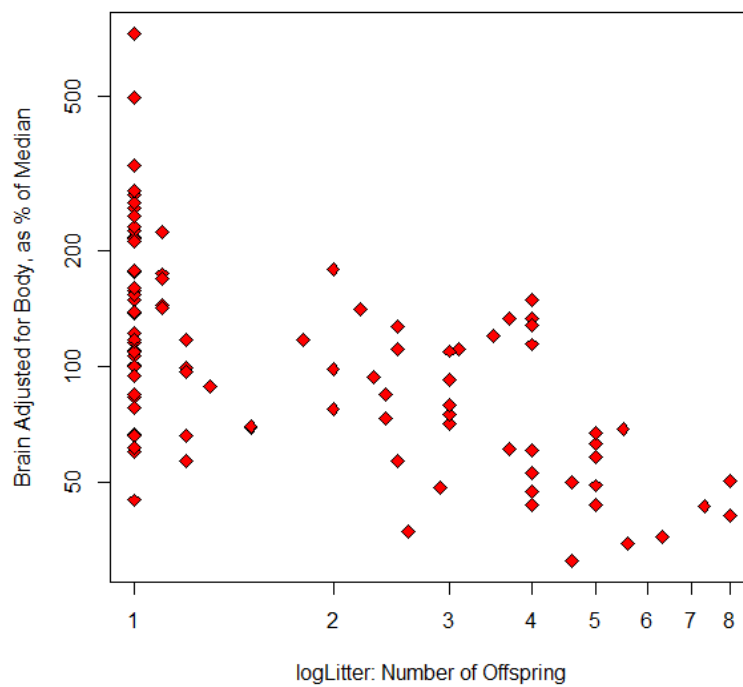


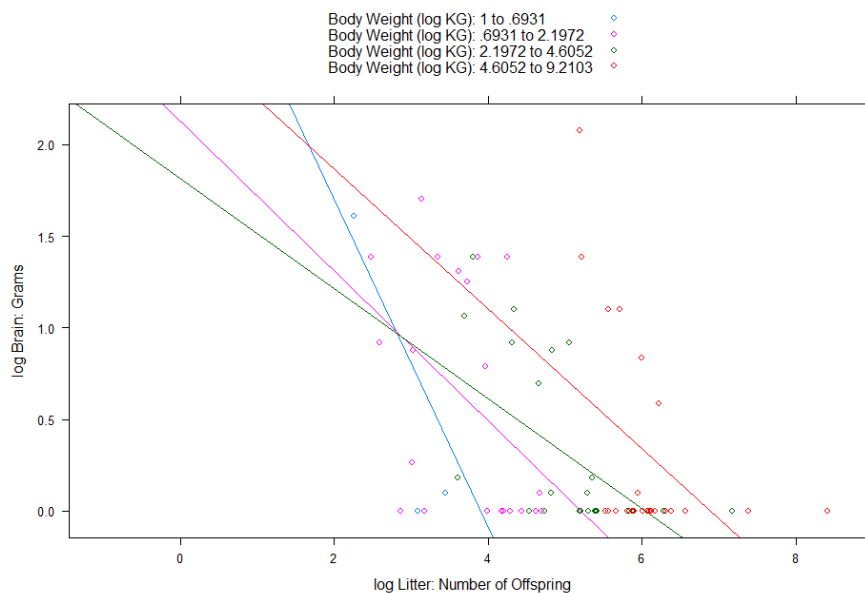
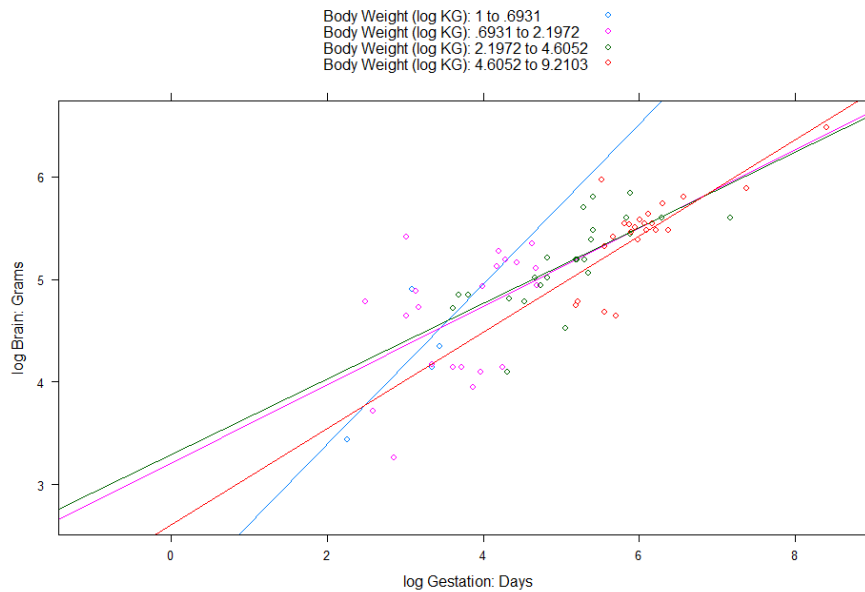
4/9/17



Component + Residual Plots



Brain Adjusted for Body vs. logGestation: 96 Mammal Species**Brain Adjusted for Body vs. logLitter: 96 Mammal Species**



The roughly parallel slopes of the lines across Body weight group confirm the known relation between Brain weight and Body size. Model-building therefore involved an assessment of various models that included and controlled for the covariate Body size. As the interaction between logGestation and logLitter proved to be statistically insignificant, the most parsimonious model for the question of interest included all three parameters (logBody, logGestation, and logLitter) without interaction.

Part B) Model – Log-Brain vs. Log-Body, Log-Gestation, and Log-Litter *without Interaction*

$$\log\text{Brain} = \beta_0 + \beta_1 \log\text{Body} + \beta_2 \log\text{Gestation} + \beta_3 \log\text{Litter} + \varepsilon$$

With ANOVA table statistics and VIF scores, the choice of this model over others was validated by comparison, which produced the following figures for the model of choice:

```
> anova(logfit_no_int)
```

Analysis of Variance Table

Response: logBrain

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
logBody	1	416	416	1847.45	< 2e-16 ***
logGestation	1	9	9	40.21	8.4e-09 ***
logLitter	1	2	2	7.15	0.0089 **
Residuals	92	21	0		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> vif(logfit_no_int)
```

	logBody	logGestation	logLitter
	3.793	6.268	2.545

Part C) Model Fit and Parameter Estimate Tables (SAS /R)

SAS

```
proc glm data = brainz plots = diagnostics(label);
model logBrain = logBody logGestation logLitter / solution;
title 'Regression of logBrain on logBody, logGestation, & logLitter';
run; quit;
```

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	0.8548219230	0.66167247	1.29	0.1996
logBody	0.5750713812	0.03258789	17.65	<.0001
logGestation	0.4179420896	0.14078249	2.97	0.0038
logLitter	-0.3100711670	0.11592709	-2.67	0.0089

R

```
> require(Sleuth3)
> require(mosaic)
> options(digits = 4)
> attach(case0902)
> case0902$logBrain <- with(case0902, log(Brain))
> case0902$logBody <- with(case0902, log(Body))
> case0902$logGestation <- with(case0902, log(Gestation))
> case0902$logLitter <- with(case0902, log(Litter))
```

```
> logfit_no_int <- lm(logBrain ~ logBody + logGestation + logLitter, data = case0902)
> summary(logfit_no_int)
```

Call:

```
lm(formula = logBrain ~ logBody + logGestation + logLitter, data = case0902)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.9541	-0.2964	-0.0311	0.2811	1.5749

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8548	0.6617	1.29	0.1996
logBody	0.5751	0.0326	17.65	<2e-16 ***
logGestation	0.4179	0.1408	2.97	0.0038 **
logLitter	-0.3101	0.1159	-2.67	0.0089 **

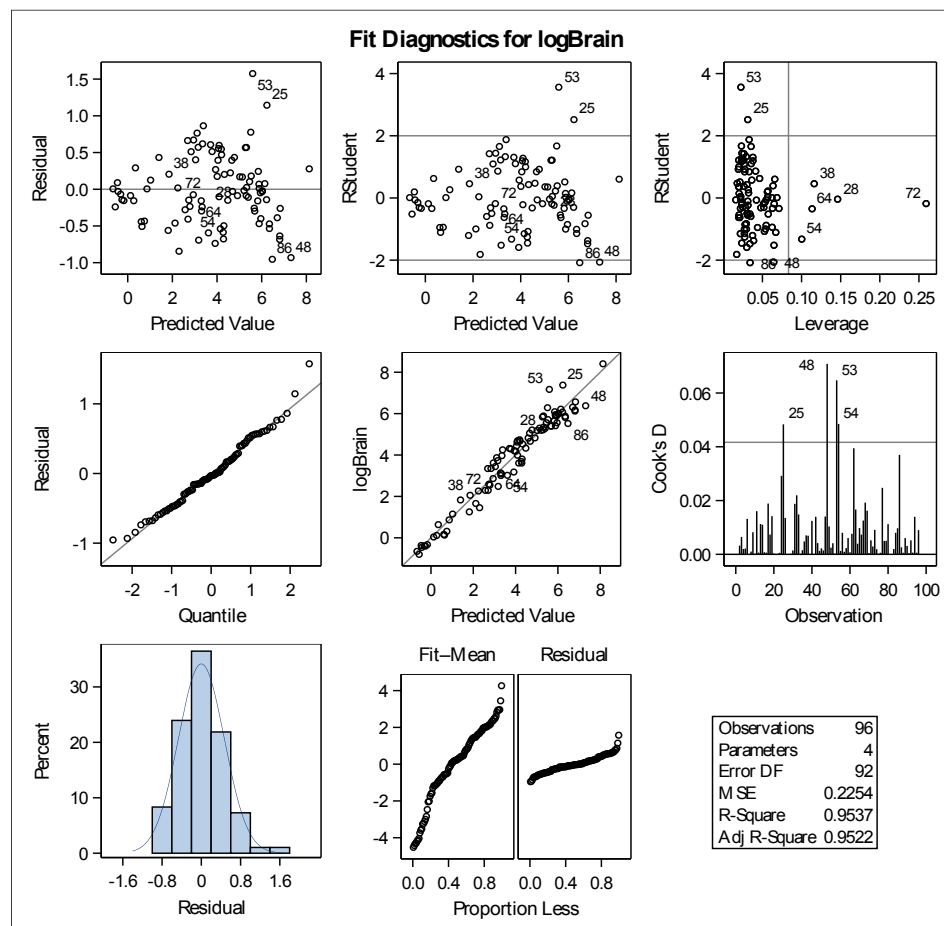
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.475 on 92 degrees of freedom

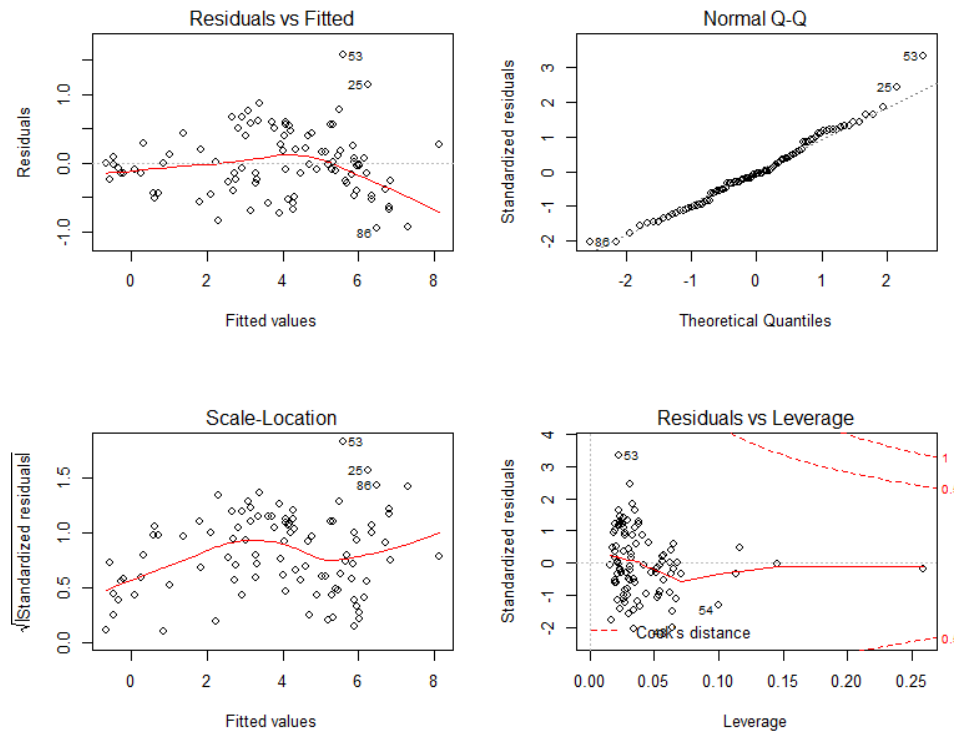
Multiple R-squared: 0.954, Adjusted R-squared: 0.952

F-statistic: 632 on 3 and 92 DF, p-value: <2e-16

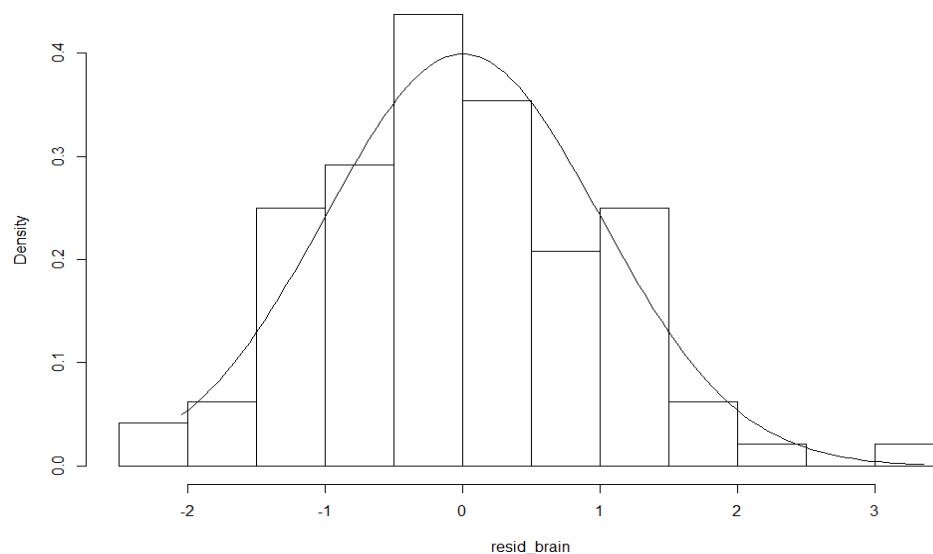
Part D) Residual Plots and Model Evaluation (SAS /R)



To assess *normality* and *constant variance* in both SAS/R, the **Residual-Predicted Value** plot adequately demonstrates random scatter about the zero line. Furthermore, observations on the **Studentized Residual-Predicted Value** plots remain largely within 2 standard deviations. **QQ-plots** exhibit adequate residual normality. Similarly, **Residual Histograms** exhibit normality under the density curve.



Distribution of Standardized Residuals: logBrain ~ logBody + logGestation + logLitter: No Interaction



Part E) Model Interpretation

To interpret the parameter coefficients, three separate linear equations were constructed:

WRT logBody:

$$\log \text{Brain} = .8548 + .5751 \log \text{Body} + .4179 \log \text{Gestation} - .3101 \log \text{Litter} + \varepsilon$$

$$= .8548 + .5751(\log \text{Body}) + .4179(0) - .3101(0) + \varepsilon$$

$$= .8548 + .5751(\log \text{Body}) + \varepsilon$$

$$\text{Brain} = e^{(.8548)} * e^{(.5751 \log \text{Body})} * \varepsilon$$

$$= (2.3509) \text{Body}^{.5751} * \varepsilon$$

→ The predicted median Brain weight of a mammal with a Body weight of 1 kg is 2.3509g. Doubling the Body weight multiplies the predicted median Brain weight by $(2^{.5751})$, or 1.4898, increasing it by 49%.

WRT logGestation:

$$\log \text{Brain} = .8548 + .5751 \log \text{Body} + .4179 \log \text{Gestation} - .3101 \log \text{Litter} + \varepsilon$$

$$= .8548 + .5751(0) + .4179(\log \text{Gestation}) - .3101(0) + \varepsilon$$

$$= .8548 + .4179(\log \text{Gestation}) + \varepsilon$$

$$\text{Brain} = e^{(.8548)} * e^{(.4179 \log \text{Gestation})} * \varepsilon$$

$$= (2.3509) \text{Gestation}^{.4179} * \varepsilon$$

→ The predicted median Brain weight of a mammal with a Gestation period of 1 day is 2.3509g. Doubling the Gestation period multiplies the predicted median Brain weight by $(2^{.4179})$, or 1.3360, increasing it by 34%.

WRT logLitter:

$$\log \text{Brain} = .8548 + .5751 \log \text{Body} + .4179 \log \text{Gestation} - .3101 \log \text{Litter} + \varepsilon$$

$$= .8548 + .5751(0) + .4179(0) - .3101(\log \text{Litter}) + \varepsilon$$

$$= .8548 - .3101(\log \text{Litter}) + \varepsilon$$

$$\text{Brain} = e^{(.8548)} * e^{(-.3101 \log \text{Litter})} * \varepsilon$$

$$= (2.3509)\text{Litter}^{-.3101} * \varepsilon$$

→ The predicted median Brain weight of a mammal with a Litter size of 1 offspring is 2.3509g. Doubling the Litter size multiplies the predicted median Brain weight by $(2^{-.3101})$, or .8066, decreasing it by 19%.

Bonus Question – Part A)

From the ANOVA tables in SAS/R, the **degrees of freedom** used to estimate the error term (**MSE**) were **df = 92**.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	427.0755192	142.3585064	631.60	<.0001
Error	92	20.7360765	0.2253921		
Corrected Total	95	447.8115958			

R-Square	Coeff Var	Root MSE	logBrain Mean
0.953695	12.28479	0.474755	3.864575

```
> anova(logfit_no_int)
Analysis of Variance Table
```

Response: logBrain

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
logBody	1	416	416	1847.45	< 2e-16 ***
logGestation	1	9	9	40.21	8.4e-09 ***
logLitter	1	2	2	7.15	0.0089 **
Residuals	92	21	0		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> summary(logfit_no_int)
```

...

Residual standard error: 0.475 on 92 degrees of freedom
Multiple R-squared: 0.954, Adjusted R-squared: 0.952
F-statistic: 632 on 3 and 92 DF, p-value: <2e-16

Bonus Question – Part B)

Similarly, from the ANOVA tables in SAS/R, the **estimate of the MSE** was .2254; the Root MSE (residual standard error) was 0.4748.

SAS Code – Question 1

```
FILENAME REFFILE '/home/jrasmusvorrath0/ex0722.csv';
```

```
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=WORK.IMPORT;  
    GETNAMES=YES;
```

```
RUN;
```

```
PROC CONTENTS DATA=WORK.IMPORT; RUN;
```

```
data crabz; set work.import;  
logForce = log(Force);  
logHeight = log(Height);  
run;
```

```
proc print data = crabz; run;
```

```
*proc means data = crabz;  
*class Species;  
*run;
```

```
ODS GRAPHICS / ATTRPRIORITY=NONE;
```

```
proc sgplot data = crabz;  
styleattrs datasymbols = (Circle Triangle Asterisk);  
scatter x = Height y = Force / group = Species;  
run;
```

```
*proc glm data = crabz plots = diagnostics(label);  
*class Species;  
*model Force = Height | Species / solution;  
*run;* quit;
```

```
*proc sgplot data = crabz;  
*styleattrs datasymbols = (Circle Triangle Asterisk);  
*scatter x = Height y = logForce / group = Species;  
*run;
```

```
proc sgplot data = crabz;  
styleattrs datasymbols = (Circle Triangle Asterisk);  
scatter x = logHeight y = logForce / group = Species;  
title 'Scatterplot logHeight on logForce by Species';  
run;
```

```
*proc glm data = crabz plots = diagnostics(label);  
*class Species;  
*model logForce = logHeight Species / solution;  
*title 'Regression of logForce on logHeight without Species Interaction';  
*run;* quit;
```

```
proc glm data = crabz plots = diagnostics(label);  
class Species;  
model logForce = logHeight | Species / solution;  
title 'Regression of logForce on logHeight with Interaction: Ref = Lophopanopeus bellus';  
run; quit;
```

```
proc glm data = crabz plots = diagnostics(label);  
class Species (REF = "Hemigrapsus nudus");  
model logForce = logHeight | Species / solution;  
title 'Regression of logForce on logHeight with Interaction: Ref = Hemigrapsus nudus';  
run; quit;
```

```
proc glm data = crabz plots = diagnostics(label);  
class Species (REF = "Cancer productus");  
model logForce = logHeight | Species / solution;  
title 'Regression of logForce on logHeight with Interaction: Ref = Cancer productus';  
run; quit;
```

```
*proc glm data = crabz plots = diagnostics(label);  
*class Species;  
*model logForce = logHeight | Species / solution noint;  
*title 'Regression of logForce on logHeight with Interaction: Ref = Intercept';  
*run;* quit;
```

```
data dummy_crabz; set crabz;  
if Species = "Hemigrapsus nudus" then d1 = 1; else d1 = 0;  
if Species = "Cancer productus" then d2 = 1; else d2 = 0;  
int1 = d1*logHeight; int2 = d2*logHeight;  
run;
```

```
proc print data = dummy_crabz; run;
```

```
*proc reg data = dummy_crabz;  
*model logForce = d1 d2 / R CLB;  
*title 'Regression of logForce on Dummy-Coded Crabs';  
*run;* quit;
```

```
*proc reg data = dummy_crabz;  
*model logForce = logHeight d1 d2 / R CLB;  
*title 'Regression of logForce on logHeight & Dummy-Coded Crabs';  
*run;* quit;
```

```
proc reg data = dummy_crabz;
model logForce = logHeight d1 d2 int1 int2 / VIF;
title 'Regression of logHeight on logForce with Dummy-Coded Interaction Terms';
run; quit;

proc means data = dummy_crabz;
var logHeight d1 d2;
run;

data center_crabz; set dummy_crabz;
cent1 = (logHeight - 2.1442590)*(d1 - 0.3684211);
cent2 = (logHeight - 2.1442590)*(d2 - 0.3157895);
run;

proc print data = center_crabz; run;

proc reg data = center_crabz;
model logForce = logHeight d1 d2 cent1 cent2 / VIF;
title 'Regression of logForce on logHeight with ~Centered~ Interaction Terms';
run; quit;

*proc glm data = crabz plots = diagnostics(label);
*class Species;
*model logForce = logHeight | Species / solution;
*title 'Regression of logForce on logHeight with Interaction: Ref = Lophopanopeus bellus';
*run;* quit;
```

SAS Code – Question 2

```
FILENAME REFFILE '/home/jrasmusvorrath0/case0902.csv';
```

```
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=WORK.IMPORT1;  
    GETNAMES=YES;  
RUN;
```

```
PROC CONTENTS DATA=WORK.IMPORT1; RUN;
```

```
data brainz; set work.import1;  
logBrain = log(Brain);  
logBody = log(Body);  
logGestation = log(Gestation);  
logLitter = log(Litter);  
run;
```

```
proc print data = brainz; run;
```

```
*proc means data = brainz;  
*var Body logBody;  
*run;
```

```
ODS GRAPHICS / ATTRPRIORITY=NONE;
```

```
*proc sgplot data = brainz;  
*styleattrs datasymbols = (Circle Triangle Asterisk Plus Star Diamond Square Hash);  
*scatter x = Body y = Brain / group = Litter;  
*run;
```

```
*proc sgplot data = brainz;  
*styleattrs datasymbols = (Circle Triangle Asterisk Plus Star Diamond Square Hash);  
*scatter x = Gestation y = Brain / group = Litter;  
*run;
```

```
proc sgscatter data = brainz;  
matrix Brain Body Gestation Litter; *alternative syntax: matrix _NUMERIC_;  
run;
```

```
*proc sgscatter data = brainz;  
*matrix logBrain logBody logGestation Litter;  
*run;
```

```
proc sgscatter data = brainz;  
matrix logBrain logBody logGestation logLitter;  
run;
```

```
proc glm data = brainz plots = diagnostics(label);  
model logBrain = logBody logGestation logLitter logGestation*logLitter / solution;  
title 'Regression of logBrain on logBody, logGestation, logLitter, with Gestation*Litter Interaction';  
run; quit;
```

```
proc glm data = brainz plots = diagnostics(label);  
model logBrain = logBody logGestation logLitter / solution;  
title 'Regression of logBrain on logBody, logGestation, & logLitter';  
run; quit;
```

```
proc glm data = brainz plots = diagnostics(label);  
model logBrain = logBody | logGestation | logLitter / solution;  
title 'Regression of logBrain on logBody, logGestation, & logLitter with Interaction Terms';  
run; quit;
```

```
proc glm data = brainz plots = diagnostics(label);  
model logBrain = logBody logGestation / solution;  
title 'Regression of logBrain on logBody & logGestation';  
run; quit;
```

```
proc glm data = brainz plots = diagnostics(label);  
model logBrain = logBody | logGestation / solution;  
title 'Regression of logBrain on logBody & logGestation, with Interaction Terms';  
run; quit;
```

```
data dummy_brainz; set brainz;  
if Body <= 2 then dBody_small = 1; else dBody_small = 0;  
if Body > 2 AND Body <= 9 then dBody_med = 1; else dBody_med = 0;  
if Body > 9 AND Body <= 100 then dBody_big = 1; else dBody_big = 0;  
if Body > 100 and Body <= 10000 then dBody_huge = 1; else dBody_huge = 0;  
int1_G = dBody_small*logGestation;  
int1_L = dBody_small*logLitter;  
int2_G = dBody_med*logGestation;  
int2_L = dBody_med*logLitter;  
int3_G = dBody_big*logGestation;  
int3_L = dBody_big*logLitter;  
int4_G = dBody_huge*logGestation;  
int4_L = dBody_huge*logLitter;  
run;
```

```
proc print data = dummy_brainz; run;
```

```
*proc reg data = dummy_brainz;  
*model logBrain = dBody_small dBody_med dBody_big dBody_huge / R CLB;  
*title 'Regression of logBrain on Dummy-Coded Body Weight';  
*run;* quit;
```

```

*proc reg data = dummy_brainz;
*model logBrain = logGestation logLitter dBody_small dBody_med dBody_big dBody_huge / R CLB;
*title 'Regression of logBrain on logGestation, logLitter, & Dummy-Coded Body Weight';
*run;* quit;

proc reg data = dummy_brainz;
model logBrain = logGestation logLitter dBody_small dBody_med dBody_big dBody_huge int1_G int1_L i
nt2_G int2_L int3_G int3_L int4_G int4_L / VIF;
title 'Regression of logBrain on logGestation, logLitter, Dummy-Coded Body Weight & Interaction Terms'
;
run; quit;

proc means data = dummy_brainz;
var logGestation logLitter dBody_small dBody_med dBody_big dBody_huge;
title 'Means of logGestation, logLitter, & Dummy-Coded Body Weight';
run;

data center_brainz; set dummy_brainz;
cent1_G = (logGestation - 4.7132378)*(dBody_small - 0.2500000);
cent1_L = (logLitter - 0.5975880)*(dBody_small - 0.2500000);
cent2_G = (logGestation - 4.7132378)*(dBody_med - 0.2500000);
cent2_L = (logLitter - 0.5975880)*(dBody_med - 0.2500000);
cent3_G = (logGestation - 4.7132378)*(dBody_big - 0.2604167);
cent3_L = (logLitter - 0.5975880)*(dBody_big - 0.2604167);
cent4_G = (logGestation - 4.7132378)*(dBody_huge - 0.2395833);
cent4_L = (logLitter - 0.5975880)*(dBody_huge - 0.2395833);
run;

proc print data = center_brainz; run;

proc reg data = center_brainz;
model logBrain = logGestation logLitter dBody_small dBody_med dBody_big dBody_huge cent1_G cent1
_L cent2_G cent2_L cent3_G cent3_L cent4_G cent4_L / VIF;
title 'Regression of logBrain on logGestation, logLitter with ~Centered~ Body Weight Interaction Terms';
run; quit;

proc reg data = center_brainz;
model logBrain = logGestation logLitter dBody_small dBody_med dBody_big dBody_huge cent1_G cent1
_L cent2_G cent2_L cent3_G cent3_L cent4_G cent4_L / noint VIF;
title 'Regression of logBrain on logGestation, logLitter with ~Centered~ Body Weight Interaction Terms';
run; quit;

*proc reg data = center_brainz;
*model logBrain = logGestation dBody_small dBody_med dBody_big dBody_huge cent1_G cent2_G cent
3_G cent4_G / VIF;
*title 'Regression of logBrain on logGestation, with ~Centered~ Body Weight Interaction Terms';
*run;* quit;

```



```
*proc glm data = brainz plots = diagnostics(label);  
*model logBrain = logBody logGestation logLitter logGestation*logLitter / solution;  
*title 'Regression of logBrain on logBody, logGestation, logLitter, with logGestation*logLitter Interaction'  
;  
*run;* quit;
```

```
*proc glm data = brainz plots = diagnostics(label);  
*model logBrain = logBody logGestation / solution;  
*title 'Regression of logBrain on logBody & logGestation';  
*run;* quit;
```

```
*proc glm data = brainz plots = diagnostics(label);  
*model logBrain = logBody logGestation logBody*logGestation / solution;  
*title 'Regression of logBrain on logBody, logGestation, & logBody*logGestation Interaction';  
*run;* quit;
```

R Code – Question 1

```

> require(Sleuth3)
> require(mosaic)
> options(digits = 4)
> attach(ex0722)

> ex0722$logHeight <- with(ex0722, log(Height))
> ex0722$logForce <- with(ex0722, log(Force))
> summary(ex0722)

> plot(logHeight, logForce, col = as.factor(Species), pch = as.numeric(as.factor(Species)))
> legend("topleft", as.vector(unique(Species)), col = as.factor(unique(Species)), pch = as.numeric(as.factor(unique(Species))))

> xyplot(logHeight ~ logForce, groups = Species, type = c("p", "r"), data = ex0722, auto.key = TRUE, ylab = "log Force: Newtons", xlab = "log Height: mm")

> par(mfrow = c(2,2))
> logfit_Default <- lm(logForce ~ logHeight*Species)
> plot(logfit_Default)
> anova(logfit_Default)
> summary(logfit_Default)

> newType <- relevel(Species, ref = "Hemigrapsus nudus")
> logfit_H <- lm(logForce ~ logHeight*newType)
> plot(logfit_H)
> anova(logfit_H)
> summary(logfit_H)

> newerType <- relevel(Species, ref = "Lophopanopeus bellus")
> logfit_L <- lm(logForce ~ logHeight*newerType)
> plot(logfit_L)
> anova(logfit_L)
> summary(logfit_L)

> par(mfrow = c(1,1))
> resid_crab <- stdres(logfit_Default)
> hist(resid_crab, freq = FALSE, main = "Distribution of Standardized Residuals: REF = C")
> xfit_C <- seq(min(resid_crab), max(resid_crab), length = 100)
> yfit_C <- dnorm(xfit_C)
> lines(xfit_C, yfit_C)

> resid_crab_H <- stdres(logfit_H)
> hist(resid_crab_H, freq = FALSE, main = "Distribution of Standardized Residuals: REF = H")
> xfit_H <- seq(min(resid_crab_H), max(resid_crab_H), length = 100)
> yfit_H <- dnorm(xfit_H)
> lines(xfit_H, yfit_H)

> resid_crab_L <- stdres(logfit_L)
> hist(resid_crab_L, freq = FALSE, main = "Distribution of Standardized Residuals: REF = L")
> xfit_L <- seq(min(resid_crab_L), max(resid_crab_L), length = 100)
> yfit_L <- dnorm(xfit_L)
> lines(xfit_L, yfit_L)

```

R Code – Question 2

```

> attach(case0902)
> case0902$logBrain <- with(case0902, log(Brain))
> case0902$logBody <- with(case0902, log(Body))
> case0902$logGestation <- with(case0902, log(Gestation))
> case0902$logLitter <- with(case0902, log(Litter))
> summary(case0902)
> panel.hist = function(x, ...){
+   usr = par("usr")
+   on.exit(par(usr))
+   par(usr = c(usr[1:2], 0, 1.5))
+   h = hist(x, plot = FALSE)
+   breaks = h$breaks
+   nB = length(breaks)
+   y = h$counts
+   y = y/max(y)
+   rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
+ }

> panel.lm = function(x, y, col = par("col"), bg = NA, pch = par("pch"), cex
= 1, col.lm = "red", ...){
+   points(x, y, pch = pch, col = col, bg = bg, cex = cex)
+   ok = is.finite(x) & is.finite(y)
+   if (any(ok))
+     abline(lm(y[ok] ~ x[ok]))
+ }

> pairs(~Brain + Body + Gestation + Litter, lower.panel = panel.smooth, diag.
panel = panel.hist, upper.panel = panel.lm, data = case0902)

> pairs(~logBrain + logBody + logGestation + logLitter, lower.panel = panel.s
mooth, diag.panel = panel.hist, upper.panel = panel.lm, data = case0902)

> xyplot(logBrain ~ jitter(logLitter), data = case0902)
> xyplot(Brain ~ jitter(Litter), scales = list(y = list(log = TRUE), x =
list(log = TRUE)), data = case0902)

> logfit_full_interact <- lm(logBrain ~ logBody * logGestation * logLitter, d
ata = case0902)
> par(mfrow = c(2,2))
> plot(logfit_full_interact)
> anova(logfit_full_interact)
> summary(logfit_full_interact)

> logfit_part_int_GL <- lm(logBrain ~ logBody + logGestation + logLitter + lo
gGestation*logLitter, data = case0902)
> par(mfrow = c(2,2))
> plot(logfit_part_int_GL)
> anova(logfit_part_int_GL)
> summary(logfit_part_int_GL)

> logfit_no_int <- lm(logBrain ~ logBody + logGestation + logLitter, data = c
ase0902)
> par(mfrow = c(2,2))
> plot(logfit_no_int)
> anova(logfit_no_int)
> summary(logfit_no_int)

```

```

> logfit_BG_no_int <- lm(logBrain ~ logBody + logGestation, data = case0902)
> par(mfrow = c(2,2))
> plot(logfit_BG_no_int)
> anova(logfit_BG_no_int)
> summary(logfit_BG_no_int)

> logfit_BG_int <- lm(logBrain ~ logBody * logGestation, data = case0902)
> par(mfrow = c(2,2))
> plot(logfit_BG_int)
> anova(logfit_BG_int)
> summary(logfit_BG_int)

> case0902$weightcut = cut(case0902$Body, breaks = c(0, 2, 9, 100, 10000), labels = c("Body weight: 0kg to 2kg", "Body weight: 2kg to 9kg", "Body weight: 9kg to 100kg", "Body weight: 100kg to 10000kg"))

> xyplot(Brain ~ jitter(Litter) | weightcut, scales = list(y = list(log = TRUE), x = list(log = TRUE)), type = c("p", "r"), data = case0902)

> log(c(0, 2, 9, 100, 10000))

> case0902$log_weightcut = cut(case0902$logBody, breaks = c(.01, .6931, 2.1972, 4.6052, 9.2103), labels = c("Body weight (log KG): .01 to .6931", "Body weight (log KG): .6931 to 2.1972", "Body weight (log KG): 2.1972 to 4.6052", "Body weight (log KG): 4.6052 to 9.2103"))
> xyplot(logBrain ~ jitter(logLitter) | log_weightcut, scales = list(y = list(log = TRUE), x = list(log = TRUE)), type = c("p", "r"), data = case0902)
> xyplot(logBrain ~ jitter(logGestation) | log_weightcut, scales = list(y = list(log = TRUE), x = list(log = TRUE)), type = c("p", "r"), data = case0902)

> xyplot(logGestation ~ logBrain, groups = log_weightcut, type = c("p", "r"), data = case0902, auto.key = TRUE, xlab = "log Gestation: Days", ylab = "log Brain: Grams")
> xyplot(logLitter ~ logBrain, groups = log_weightcut, type = c("p", "r"), data = case0902, auto.key = TRUE, xlab = "log Litter: Number of Offspring", ylab = "log Brain: Grams")

> newCut <- relevel(log_weightcut, ref = "Body weight (log KG): 4.6052 to 9.2103")
> logfit_WG_huge <- lm(logBrain ~ logGestation * newCut, data = case0902)
> par(mfrow = c(2,2))
> plot(logfit_WG_huge)
> anova(logfit_WG_huge)
> summary(logfit_WG_huge)

> vif(logfit_full_interact)
> vif(logfit_part_int_GL)
> vif(logfit_BG_int)
> vif(logfit_WG_huge)
> vif(logfit_no_int)

> require(car)
> crPlots(logfit_no_int)

```

```
> beta_BG_no_int <- logfit_BG_no_int$coef
> logBrainAdjB_G <- logBrain - beta_BG_no_int[2]*logBody
> yAdjB_G <- exp(logBrainAdjB_G)
> ymodB_G <- 100*yAdjB_G/median(yAdjB_G)
> plot(ymodB_G ~ Gestation, log = "xy", xlab = "logGestation: Days", ylab = "
Brain Adjusted for Body, as % of Median", main = "Brain Adjusted fpr Body vs.
logGestation: 96 Mammal Species", pch = 22, bg = "blue", cex = 1.1)

> logfit_BL_no_int <- lm(logBrain ~ logBody + logLitter)
> beta_BL_no_int <- logfit_BL_no_int$coef
> logBrainAdjB_L <- logBrain - beta_BL_no_int[2]*logBody
> yAdjB_L <- exp(logBrainAdjB_L)
> ymodB_L <- 100*yAdjB_L/median(yAdjB_L)
> plot(ymodB_L ~ Litter, log = "xy", xlab = "logLitter: Number of Offspring",
ylab = "Brain Adjusted for Body, as % of Median", main = "Brain Adjusted for
Body vs. logLitter: 96 Mammal Species", pch = 23, bg = "red", cex = 1.1)

> par(mfrow = c(1,1))
> resid_brain <- stdres(logfit_no_int)
> hist(resid_brain, freq = FALSE, main = "Distribution of Standardized Residu
als: logBrain ~ logBody + logGestation + logLitter: No Interaction")
> xfit_B <- seq(min(resid_brain), max(resid_brain), length = 100)
> yfit_B <- dnorm(xfit_B)
> lines(xfit_B, yfit_B)
```