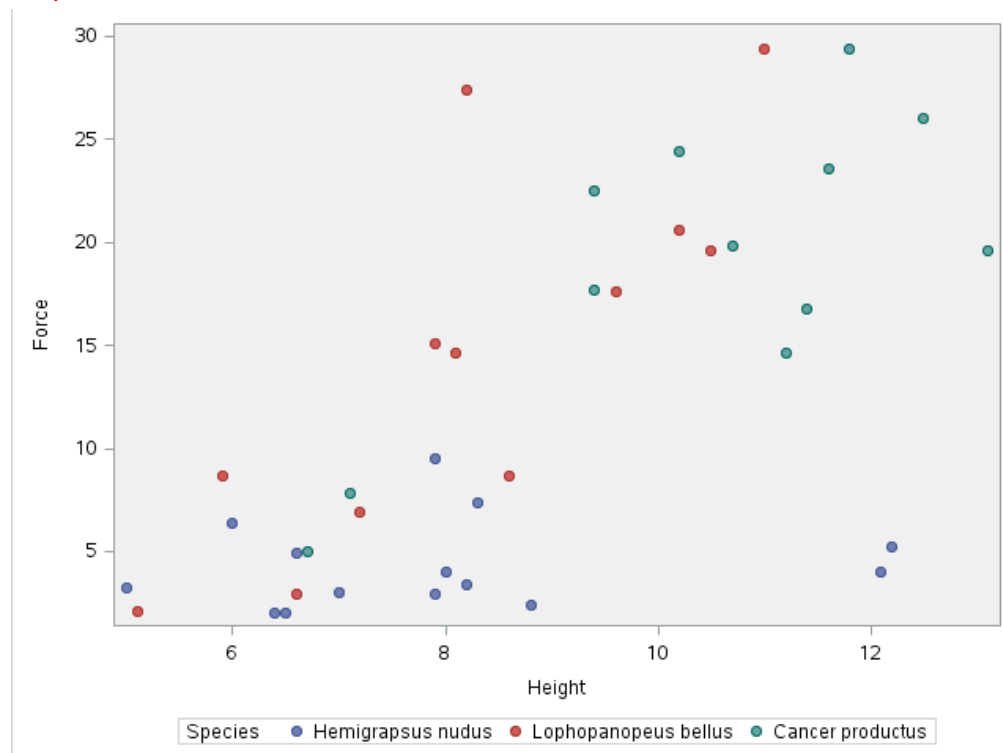


UNIT 12 HW

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.
- a. 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.

Unlogged

```
proc sgplot data=Crab;  
styleattrs wallcolor=CXF0F0F0;  
scatter x=Height y=Force / group=Species  
markerattrs=(symbol=CircleFilled size=7)  
filledoutlinedmarkers;  
run;
```



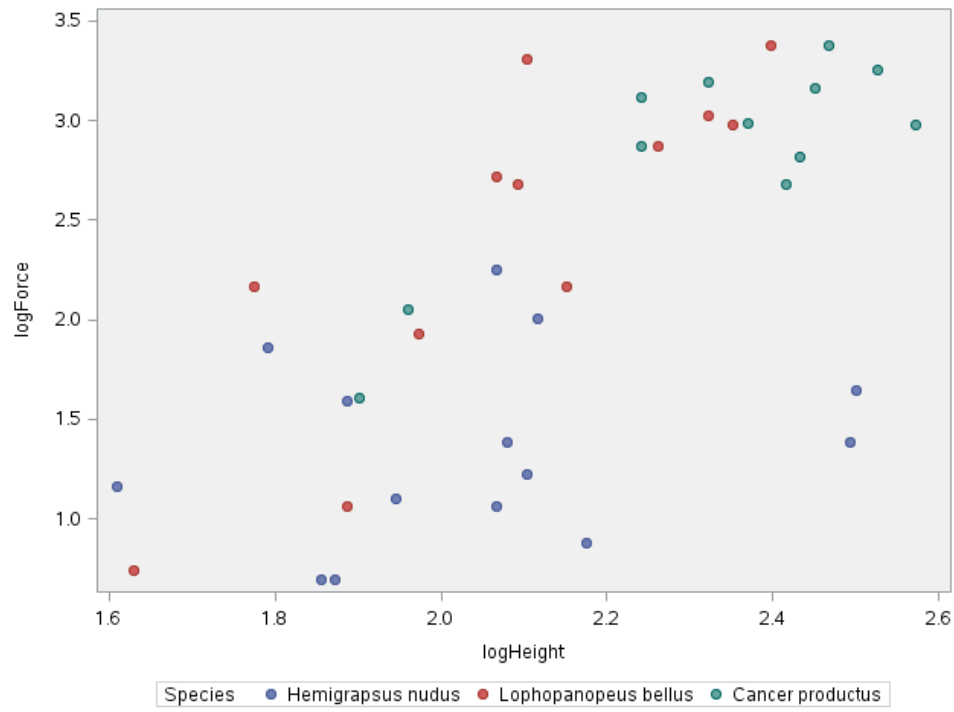
Logged

```
proc sgplot data=logCrab;
```

```

styleattrs wallcolor=CXF0F0F0;
scatter x=logHeight y=logForce/ group=Species
markerattrs=(symbol=CircleFilled size=7)
filledoutlinedmarkers;
run;

```



- b. 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.

$$\hat{\mu}_{Force|Species, Height} = \beta_0 + \beta_1 Height + \beta_2 Species + \beta_3 Height \times Species$$

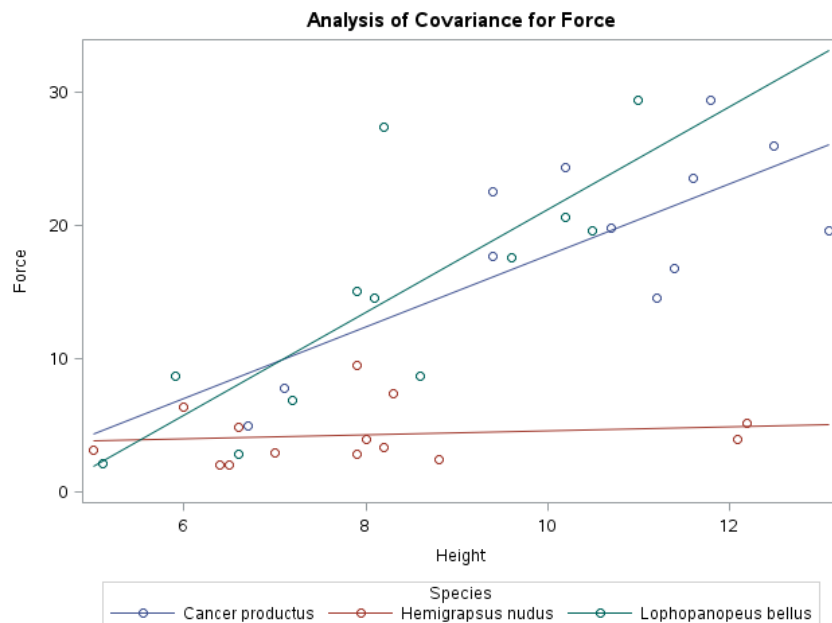
- c. Step 3: Fit the model. (Fill in the relevant betas in your equation for step 2.) Make sure you provide relevant code and the table of parameter estimates as well.

```

proc glm data=Crab plots = all;
class Species(ref = "Lophopanopeus bellus");
model Force = Height | Species / solution clparm;
run
run;

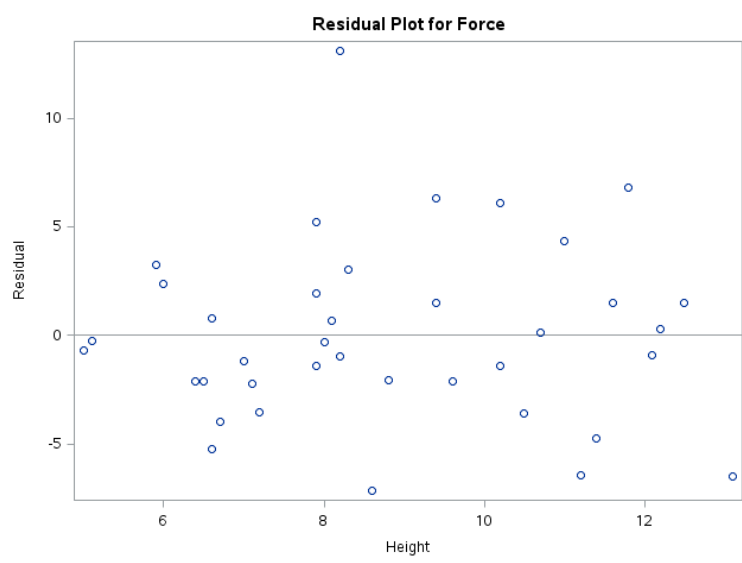
```

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-17.25249079	B	6.09642499	-2.83	0.0080
Height	3.84863387	B	0.72313494	5.32	<.0001
Species Cancer productus	8.24823996	B	9.39410117	0.88	0.3865
Species Hemigrapsus nudus	20.41259323	B	7.79426488	2.62	0.0134
Species Lophopanopeus bellus	0.00000000	B	-	-	-
Height*Species Cancer productus	-1.16876968	B	0.98883681	-1.18	0.2459
Height*Species Hemigrapsus nudus	-3.70383165	B	0.93609777	-3.96	0.0004
Height*Species Lophopanopeus bellus	0.00000000	B	-	-	-



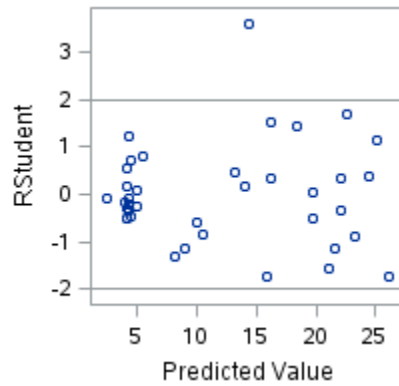
- d. Step 4: Provide a residual plot, studentized residual plot, histogram of residuals, and q-q plot of residuals to provide evidence of the appropriateness of the model. Provide a short one- or two-sentence discussion of EACH plot.

Residual Plot

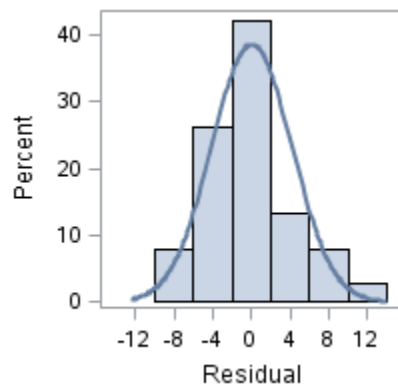


The data is spread out. The residuals look equally spread out in a random cloud.

Studentized Residual plot

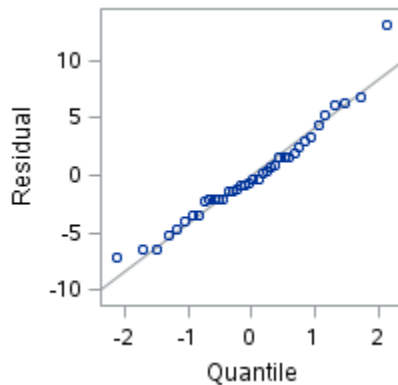


Histogram of residuals



The residuals look more normally distributed.

q-q plot of residuals



The residuals correlate almost directly with normal quantiles signs of a normal spread.

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

Since the model is sufficient

The lopho crab with a height of 0 it would only exert -17.25 units of force which makes little sense physically. A Cancer crab would be able to exert 8.25 units of force more than the Lopho crab. The Hemigraspus crab would be able to exert 20.41 more units of force than the Lopho crab.

For every single unit of prodopus height in a Lopho crab it gains 3.85 units of force, the cancer it will gain 1.17 less units of force than the lopho crab. The Hemigraspus gains 3.7 less units of force than the lopho crab.

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

$$\mu\{Force | Height, Lopho\} = -17.25 + 3.85 \text{ Height}$$

$$\mu\{Force | Height, Cancer\} = -9 + 2.68 \text{ Height}$$

$$\mu\{Force | Height, Hemi\} = 3.6 + 0.15 \text{ Height}$$

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.

```
Brain <- read.csv('/Brain.csv')
```

```
Brain =transform(Brain, logBrain=log(Brain))
```

```
Brain =transform(Brain, logBody=log(Body))
```

```
Brain =transform(Brain, logGest=log(Gestation))
```

```
Brain =transform(Brain, logLit=log(Litter))
```

```
Brain$bodyGroup <- cut(Brain$logBody, c(-Inf, -2, 0, 2, 4, 6, Inf))
```

```
levels(Brain$bodyGroup) <- c("<0.14Grams", "0.14-1Grams", "1-7.39Grams", "7.39-54.6Grams", "54.6-403Grams", ">403Grams")
```

```
Brain.lm <-lm(logBrain ~ logGest + logLit + logBody, data=Brain)
```

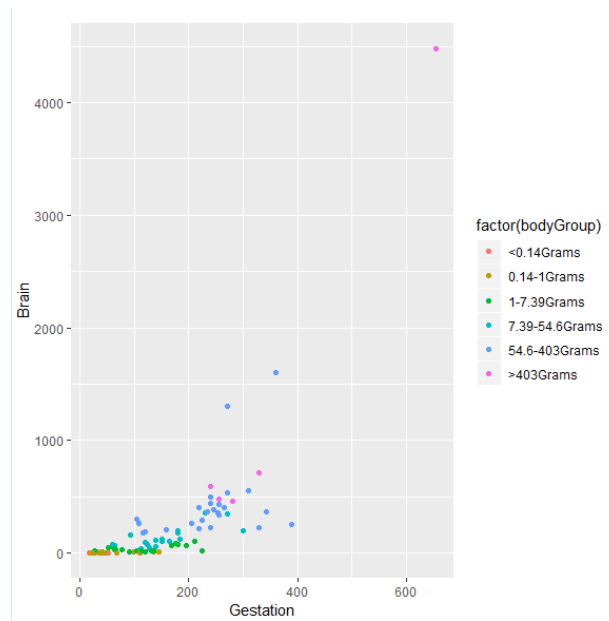
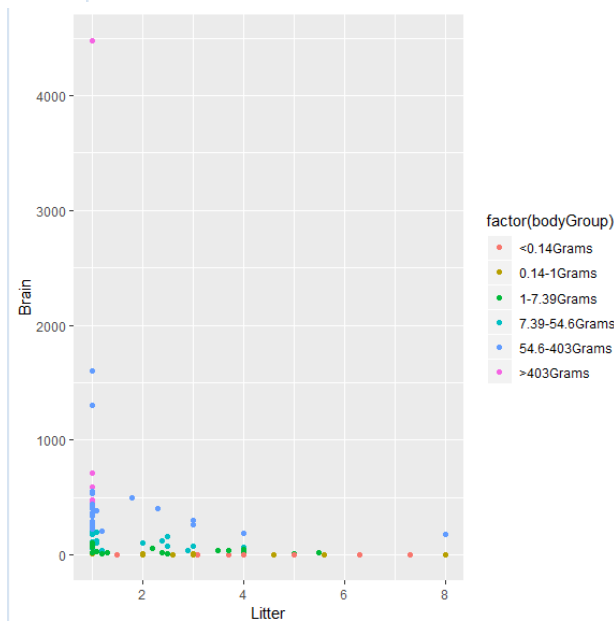
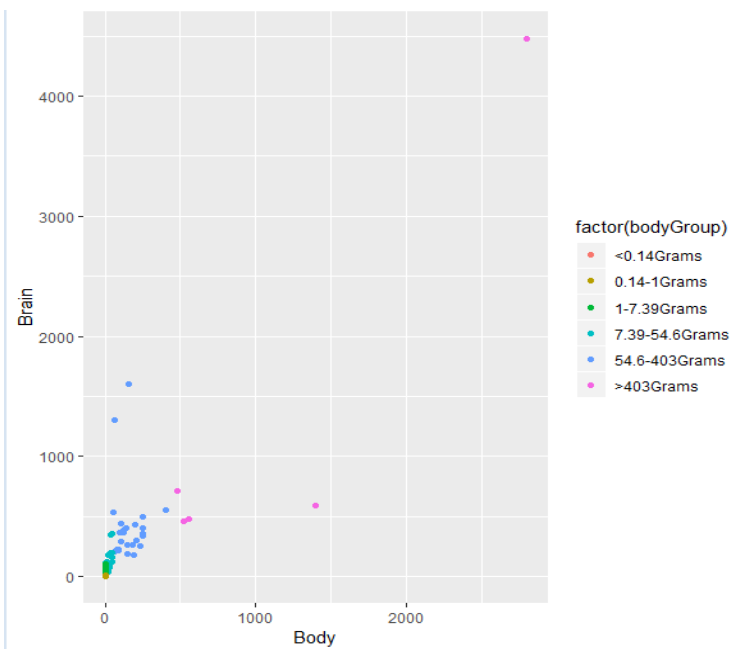
```
library(ggplot2)
```

```
ggplot(Brain, aes(Gestation, Brain))+  
geom_point(aes(color = factor(bodyGroup)))
```

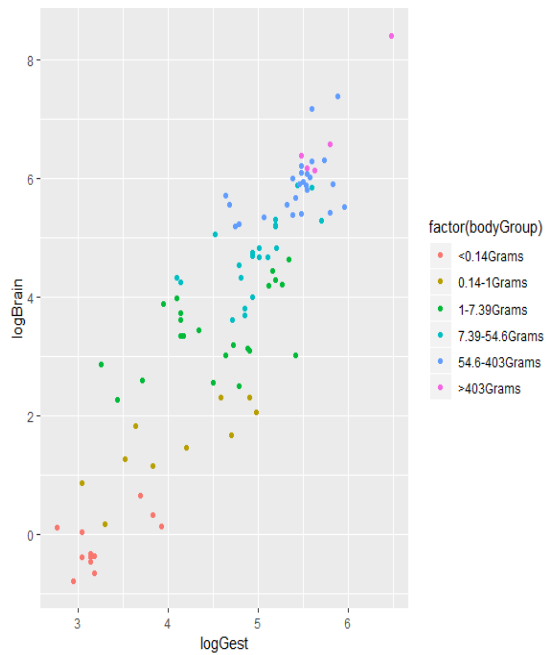
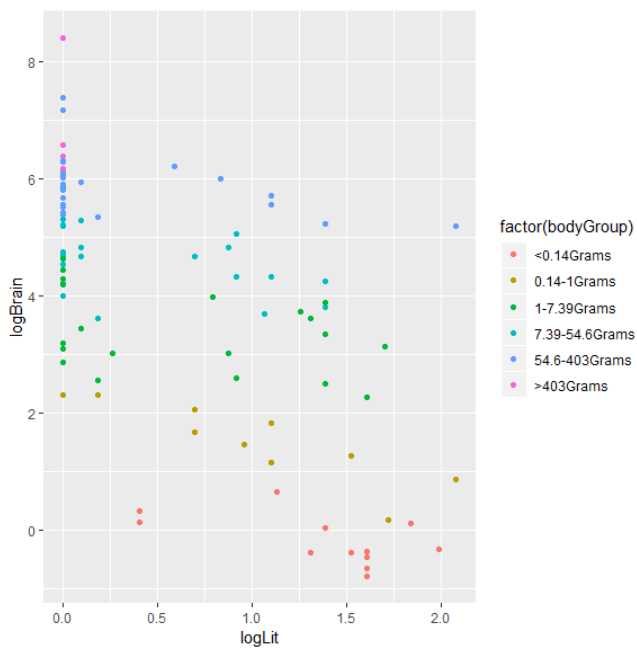
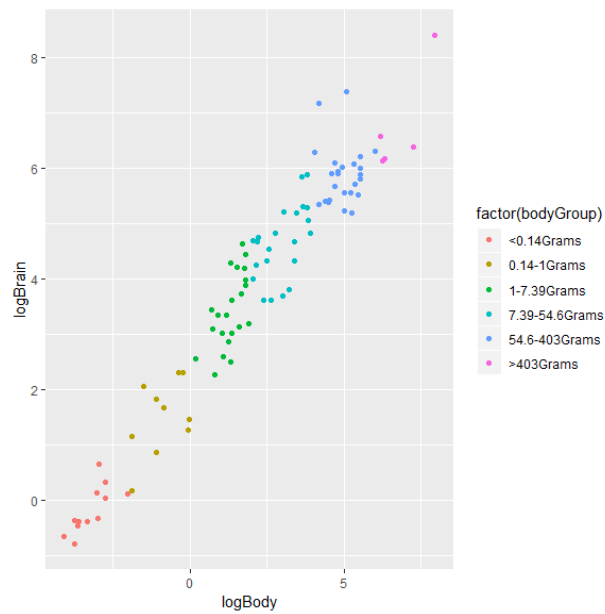
just change ggplot for each log and unlog plot

a. Step1:

Unlogged Plots



Logged Plots



b. Step 2: $\mu\{\text{brain size} \mid \text{gestation, litter, body}\} = \beta_0 + \beta_1 \text{ gestation} + \beta_2 \text{ litter} + \beta_3 \text{ body}$

c. Step 3:

`summary(Brain.lm)`

```

Call:
lm(formula = logBrain ~ logGest + logLit + logBody, data = Brain)

Residuals:
    Min       1Q   Median       3Q      Max
-0.95415 -0.29639 -0.03105  0.28111  1.57491

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.85482     0.66167   1.292  0.19962
logGest      0.41794     0.14078   2.969  0.00381 **
logLit      -0.31007     0.11593  -2.675  0.00885 **
logBody      0.57507     0.03259  17.647 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4748 on 92 degrees of freedom
Multiple R-squared:  0.9537,    Adjusted R-squared:  0.9522
F-statistic: 631.6 on 3 and 92 DF,  p-value: < 2.2e-16

```

d. Step 4:

#Store residuals from linear model

Brainres <- resid(Brain.lm)

par(mfrow=c(2,2))

#Get scatterplot of residuals

plot(Brain\$logGest,Brainres, ylab="Residuals", xlab="logGestation Period", main="Residuals versus Gestation Period")

abline (0,0)

#Get scatterplot of residuals

plot(Brain\$logLit,Brainres, ylab="Residuals", xlab="logLitter Size", main="Residuals versus Litter Size")

abline (0,0)

#Get scatterplot of residuals

plot(Brain\$logBody, Brainres, ylab="Residuals", xlab="logBody Size", main="Residuals versus Body Size")

abline (0,0)


```
##Store studentized residuals
```

```
studresBrain <- rstudent(Brain.lm)
```

```
##Histogram
```

```
hist(studresBrain, freq=FALSE, main="Distribution of Studentized Residuals",
```

```
xlab="Studentized Residuals", ylab="Density", ylim=c(0,0.5))
```

```
##Create range of x-values for normal curve
```

```
xfit <- seq(min(studresBrain)-1, max(studresBrain)+1, length=40)
```

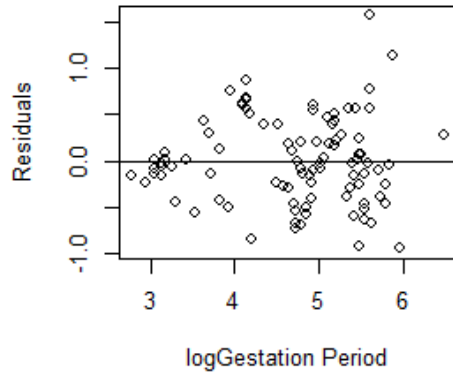
```
##Generate values from the normal distribution at the specified values
```

```
yfit <- (dnorm(xfit))
```

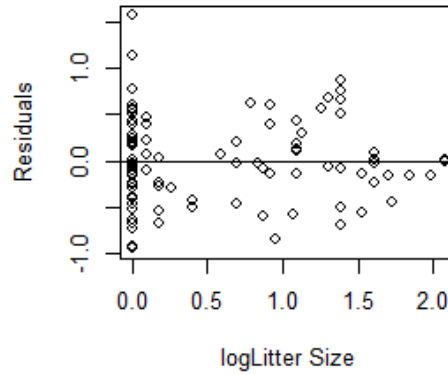
```
##Add the normal curve
```

```
lines(xfit, yfit, ylim=c(0,0.5))
```

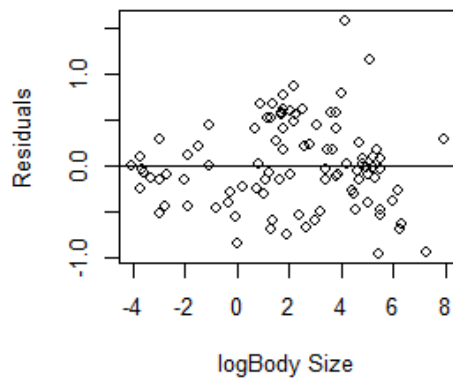
Residuals versus Gestation Period



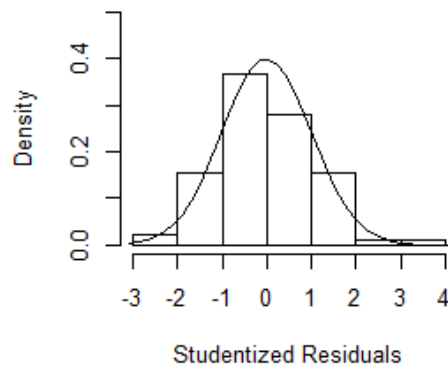
Residuals versus Litter Size



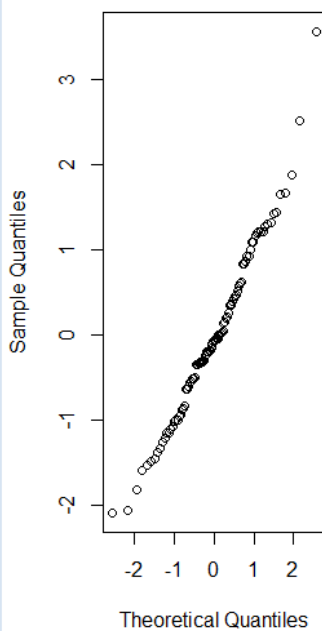
Residuals versus Body Size



Distribution of Studentized Residuals



Normal Q-Q Plot



Linearity: Met with log-log model. (view scatterplots)

Normality: Log-log model looks better. (view histograms)

Equal standard deviations: Log-log model looks much better. (view residual scatter plots)

Independence: We will assume independence

We will proceed to make inferences on a log-log model.

e. Step 5

If gestation goes up by 1 percent then brain size goes up by .41% but if litter goes up 1 percent then brain size goes down .31%. So both estimates are significant when accounting for body size also (brain size increases by .57%).

d. Bonus

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

32

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

0.432909

c. Repeat 1(a) in R.

d. Repeat 1(c) in R.

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

92

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

0.4748