Takehome 2

On my honor, I have not had any form of communication about this exam with any other individual(including other students, teaching assistants, instructors, etc.) -Dandong Tu

1a
Based on the summary of **m1** attached in supporting materal **1a**, the expected response is significantly different among the levels since the **p-value** are both small when we move from base level to other two levels.

```
data1=read.table("/Users/dandongtu/Desktop/takehome2.txt")
m1=lm(data1$Energy~data1$Type)
```

1b

```
Anova(m1)
```

```
## Anova Table (Type II tests)
##
## Response: data1$Energy
##             Sum Sq Df F value    Pr(>F)
## data1$Type 17.845  2  12.504 0.0004576 ***
## Residuals  12.130 17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observed the **p-value=0.0004576**, thus we reject the $H_0 : E(Energy|X) = \beta_0$ and conclude that **Type** does have a statistically significant influence on **Energy**.

2a From the plots in supporting material **2a**, it is easy to see that the "log" transformation in **Mass** is more reasonable as a linear relation for **Energy**.

```
m2.a=lm(Energy~poly(log(Mass),2),data=data1)
```

From summary of model **m2.a** in supporting materal **2a**, we obtained **p-value=3.02e-16** for the 1st degree of polynomial,means that it's statisticlly significant. Since the **p-value** for 2nd degree of polynomial is **0.278**, means it's not statisticlly significant, so that we conclude we should not include second degree of polynomial.

2b

```
m2=lm(Energy~log(Mass),data=data1)
Anova(m2)
```

```
## Anova Table (Type II tests)
##
## Response: Energy
##             Sum Sq Df F value    Pr(>F)
## log(Mass) 29.3919  1  907.64 < 2.2e-16 ***
## Residuals  0.5829 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The **p-value=2.2e-16**, thus we reject the $H_0 : E(Energy|X) = \beta_0$ and conclude that log(**Mass**) does have a statistically significant influence on **Energy**.

3a

```
m3=lm(Energy~log(Mass)+Type,data=data1)
m3.a=lm(Energy~log(Mass)*Type,data=data1)
anova(m3,m3.a)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: Energy ~ log(Mass) + Type
## Model 2: Energy ~ log(Mass) * Type
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     16 0.55332
## 2     14 0.50487  2   0.04845 0.6718 0.5265
```

We obtained the **p-value=0.5265** from the anova table, and we fail to reject the reduced model (reduced model is adequate). Therefore, we would like to use a model without the interactions.

3b

```
Anova(m3)
```

```
## Anova Table (Type II tests)
##
## Response: Energy
##             Sum Sq Df  F value    Pr(>F)
## log(Mass) 11.5770  1 334.7662 3.758e-12 ***
## Type       0.0296  2   0.4276    0.6593
## Residuals  0.5533 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The **p-value** for **Type** is **0.6593** and so we fail to reject the null which means **Type** does not have a significant effect after adjusting for log(**Mass**). Thus, we choose the model only contains log(**Mass**).

4a

```
m4=lm(Energy~log(Mass),data=data1)
ncvTest(m4)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.6779911    Df = 1     p = 0.4102793
```

```
ncvTest(m4,~I(log(Mass)^2))
```

```
## Non-constant Variance Score Test
## Variance formula: ~ I(log(Mass)^2)
## Chisquare = 0.6283434    Df = 1     p = 0.4279636
```

The residualPlot in supporting materal **4a** does not show a clear trend of residuals.Also, based on the ncvTest, we observed that both tests have high **p-value**(**0.4103** and **0.428**). Thus, we do not have enough evidence to reject the null(the Variance is constant), and we assume the variance is constant.

4b

```
confint(m4,level = 0.98)[2,]
```

```
##       1 %      99 %
## 0.7401039 0.8771156
```

```
sandwich_se=sqrt(diag(hccm(m4,type = "hc3")))
(sandwich_ci98=c(coef(m4)-qt(0.99,20-2)*sandwich_se,coef(m4)+qt(0.99,20-2)*sandwich_se))
```

```
## (Intercept)   log(Mass) (Intercept)   log(Mass)
##  -1.7776450   0.7440359  -1.1588719   0.8731836
```

From the results, it is easy to see that the 98% CI of OLS corrected with sandwich estimator is narrower than just using OLS since the standard error is adjusted to be smaller with sandwich estimator

Supporting material:

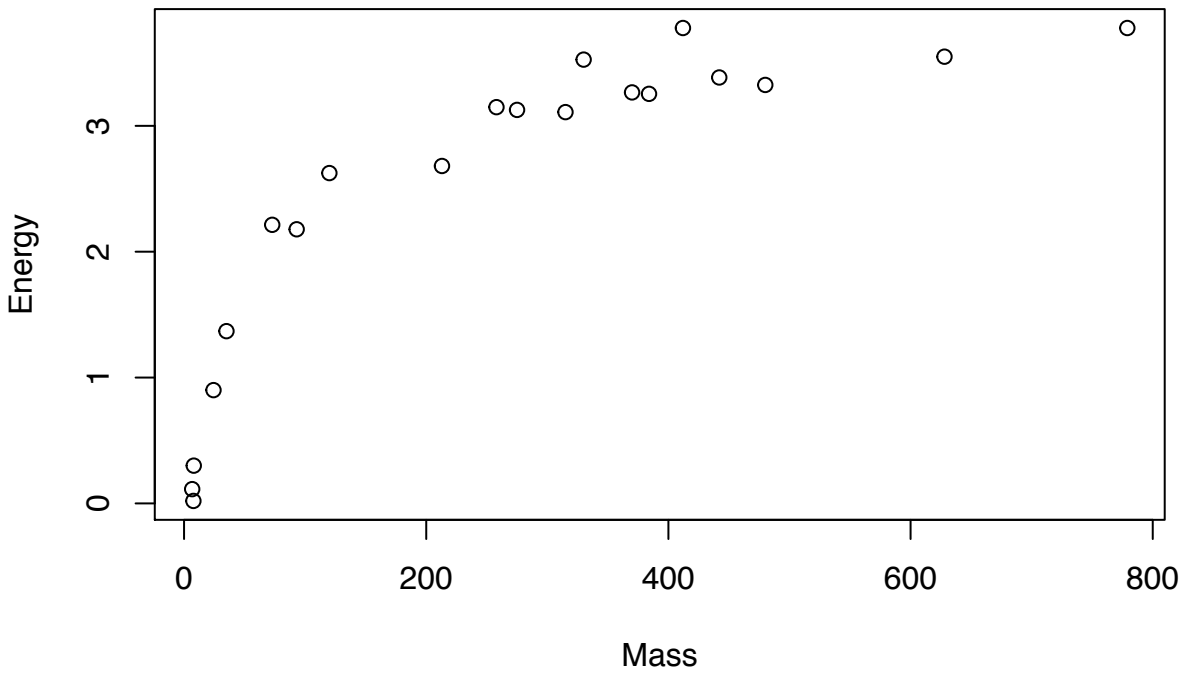
supporting materal for 1a.

```
summary(m1)
```

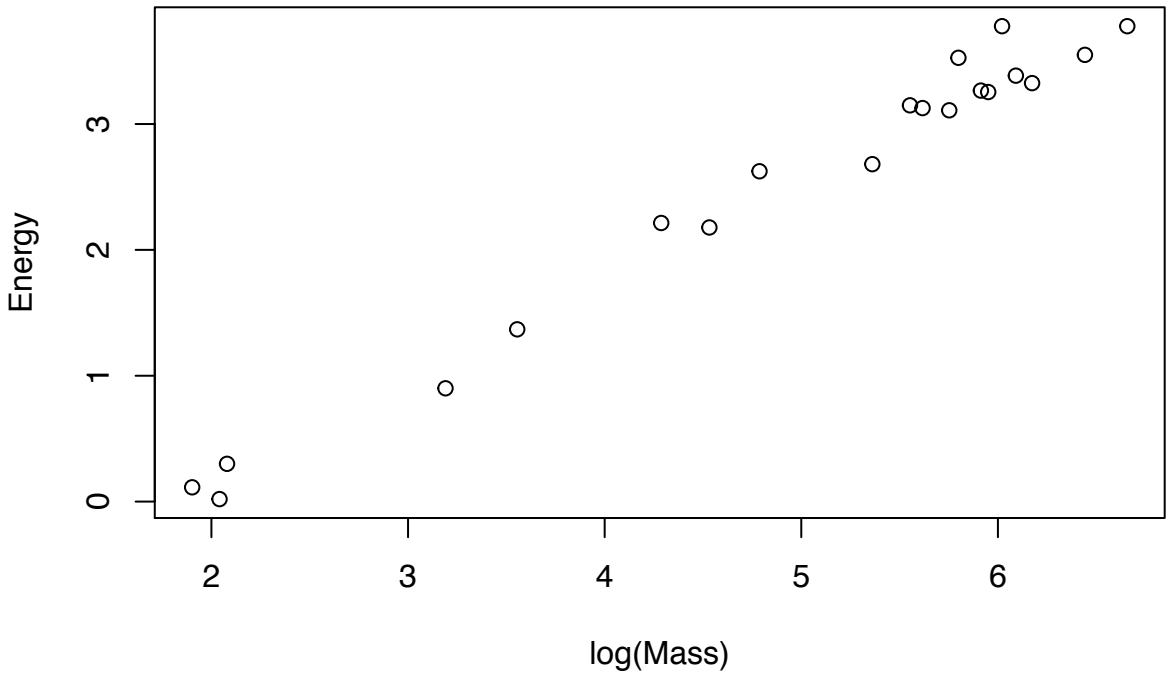
```
##
## Call:
## lm(formula = data1$Energy ~ data1$Type)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.88718 -0.39944  0.02359  0.49323  1.52531
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.6528     0.4224   1.546 0.140585
## data1$Typenon-echolocating bats   2.7433     0.5973   4.593 0.000259 ***
## data1$Typenon-echolocating birds  2.1345     0.4877   4.377 0.000411 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8447 on 17 degrees of freedom
## Multiple R-squared:  0.5953, Adjusted R-squared:  0.5477
## F-statistic:  12.5 on 2 and 17 DF,  p-value: 0.0004576
```

supporting materal for 2a.

```
plot(Energy~Mass,data=data1)
```



```r
plot(Energy~log(Mass),data=data1)
```




```r
summary(m2.a)
```

```
##
## Call:
## lm(formula = Energy ~ poly(log(Mass), 2), data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25245 -0.09864 -0.05440  0.11655  0.39090
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.48220    0.03995  62.126  < 2e-16 ***
## poly(log(Mass), 2)1   5.42143    0.17868  30.341 3.02e-16 ***
## poly(log(Mass), 2)2 -0.20033    0.17868  -1.121    0.278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1787 on 17 degrees of freedom
## Multiple R-squared:  0.9819, Adjusted R-squared:  0.9798
## F-statistic: 460.9 on 2 and 17 DF,  p-value: 1.555e-15
```

supporting materal for 3b

```r
summary(m3)
```

```
##
## Call:
## lm(formula = Energy ~ log(Mass) + Type, data = data1)
```

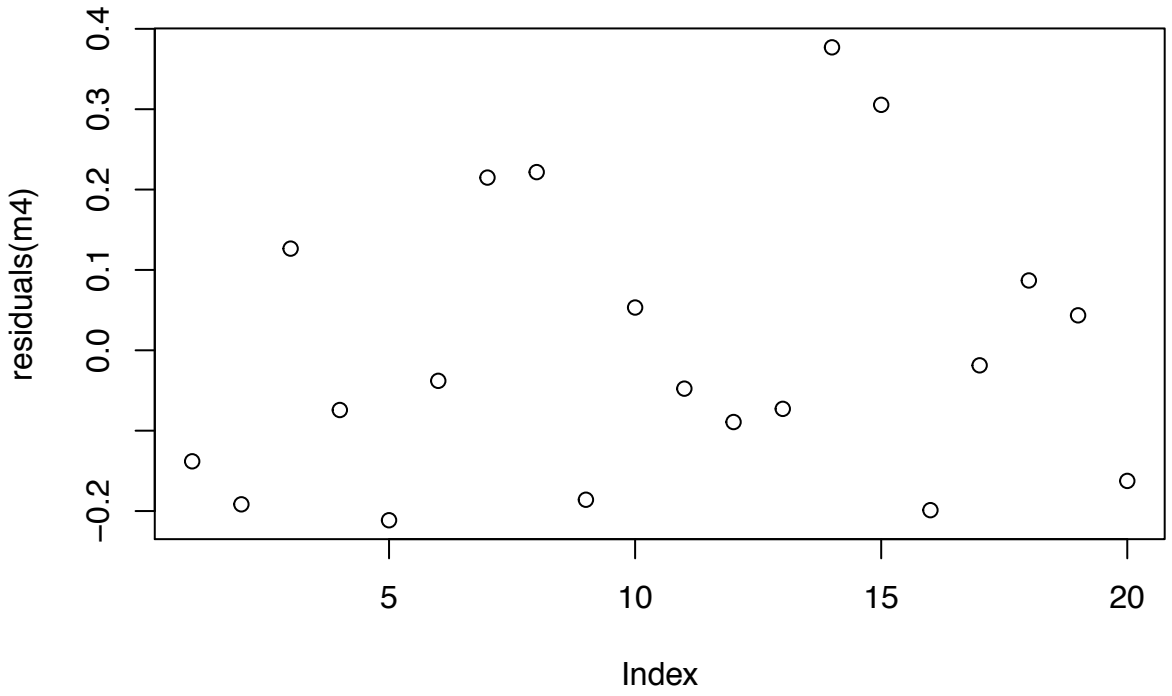
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23224 -0.12199 -0.03637  0.12574  0.34457
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.49770    0.14987  -9.993 2.77e-08 ***
## log(Mass)                   0.81496    0.04454  18.297 3.76e-12 ***
## Typenon-echolocating bats  -0.07866    0.20268  -0.388    0.703
## Typenon-echolocating birds  0.02360    0.15760   0.150    0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 16 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9781
## F-statistic: 283.6 on 3 and 16 DF,  p-value: 4.464e-14
```

supporting materal for 4a

```
summary(m4)
```

```
##
## Call:
## lm(formula = Energy ~ log(Mass), data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21143 -0.14422 -0.04284  0.09681  0.37695
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.46826    0.13716  -10.71  3.1e-09 ***
## log(Mass)    0.80861    0.02684   30.13  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.18 on 18 degrees of freedom
## Multiple R-squared:  0.9806, Adjusted R-squared:  0.9795
## F-statistic: 907.6 on 1 and 18 DF,  p-value: < 2.2e-16
```

```
plot(residuals(m4))
```


residuals(m4)
−0.2
0.0
0.1
0.2
0.3
0.4
5
10
15
20
Index