Comparison of Models: Analysis of Variance

```
> bat<-read.table("/Users/emmanuel/Data/bat.dat", header = T)
> bat
    mass type energy
1  779.0    1  43.70
2  628.0    1  34.80
3  258.0    1  23.30
4  315.0    1  22.40
5   24.3    2   2.46
6   35.0    2   3.93
7   72.8    2   9.15
8  120.0    2  13.80
9  213.0    2  14.60
10 275.0    2  22.80
11 370.0    2  26.20
12 384.0    2  25.90
13 442.0    2  29.50
14 412.0    2  43.70
15 330.0    2  34.00
16 480.0    2  27.80
17  93.0    3   8.83
18   8.0    3   1.35
19   6.7    3   1.12
20   7.7    3   1.02

> # Take Logarithms
> lmass <- log(mass)
> lenergy <- log(energy)
```

Visualize data

```
> postscript("anova1.ps")
> plot(lmass,lenergy, type = "n")
> points(lmass[type ==1],lenergy[type ==1])
> points(lmass[type ==2],lenergy[type ==2], type = "p", pch = 3)
> points(lmass[type ==3],lenergy[type ==3], type = "p", pch = 8)
```

Create dummy variables

```
> bird <- rep(0,20)
```

```
> bird[bird.index] = 1
> ebat <- rep(0,20)
> ebat[ebat.index] = 1
```

Fit multiple linear regression model and visualize the fit

```
> full.model <- lm(lenergy ~ lmass + bird + ebat)
> full.model

Call:
lm(formula = lenergy ~ lmass + bird + ebat)

Coefficients:
(Intercept)        lmass         bird         ebat
   -1.57636      0.81496      0.10226      0.07866
```

```
> postscript("anova2.ps")
> plot(lmass,lenergy, type = "n")
> points(lmass[type ==1],lenergy[type ==1])
> points(lmass[type ==2],lenergy[type ==2], type = "p", pch = 3)
> points(lmass[type ==3],lenergy[type ==3], type = "p", pch = 8)

> abline(beta[1],beta[2])
> abline(beta[1] + beta[3],beta[2])
> abline(beta[1] + beta[4],beta[2])
```

Fit a single linear regression model with `lmass` as predictor and compare this simpler model with the full model.

```
> small.model<-lm(lenergy ~ lmass)
> small.model

Call:
lm(formula = lenergy ~ lmass)

Coefficients:
(Intercept)        lmass
    -1.4683       0.8086

> anova(small.model,full.model)
Analysis of Variance Table
```

```
Model 1: lenergy ~ lmass
Model 2: lenergy ~ lmass + bird + ebat
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     18 0.58289
2     16 0.55332  2   0.02957 0.4276 0.6593
```

Removing one predictor is the same as a t-test.

```
> summary(full.model)

Call:
lm(formula = lenergy ~ lmass + bird + ebat)

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) -1.57636    0.28724  -5.488 4.96e-05 ***
lmass        0.81496    0.04454  18.297 3.76e-12 ***
bird         0.10226    0.11418   0.896    0.384
ebat         0.07866    0.20268   0.388    0.703
---
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

> full.model.minus.ebat <- update(full.model, ~. - ebat)
> anova(full.model.minus.ebat,full.model)
Analysis of Variance Table

Model 1: lenergy ~ lmass + bird
Model 2: lenergy ~ lmass + bird + ebat
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     17 0.55853
2     16 0.55332  1   0.00521 0.1506  0.703

> full.model.minus.bird <- update(full.model, ~. - bird)
> anova(full.model.minus.bird,full.model)
Analysis of Variance Table

Model 1: lenergy ~ lmass + ebat
Model 2: lenergy ~ lmass + bird + ebat
```

```
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     17 0.58106
2     16 0.55332  1   0.02774 0.8021 0.3837

\end{verbatime}

Sequential Analysis of variance

\begin{verbatim}

> anova(full.model)
Analysis of Variance Table

Response: lenergy
          Df  Sum Sq Mean Sq  F value    Pr(>F)
lmass      1 29.3919 29.3919 849.9108 2.691e-15 ***
bird       1  0.0244  0.0244   0.7045    0.4136
ebat       1  0.0052  0.0052   0.1506    0.7030
Residuals 16  0.5533  0.0346
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

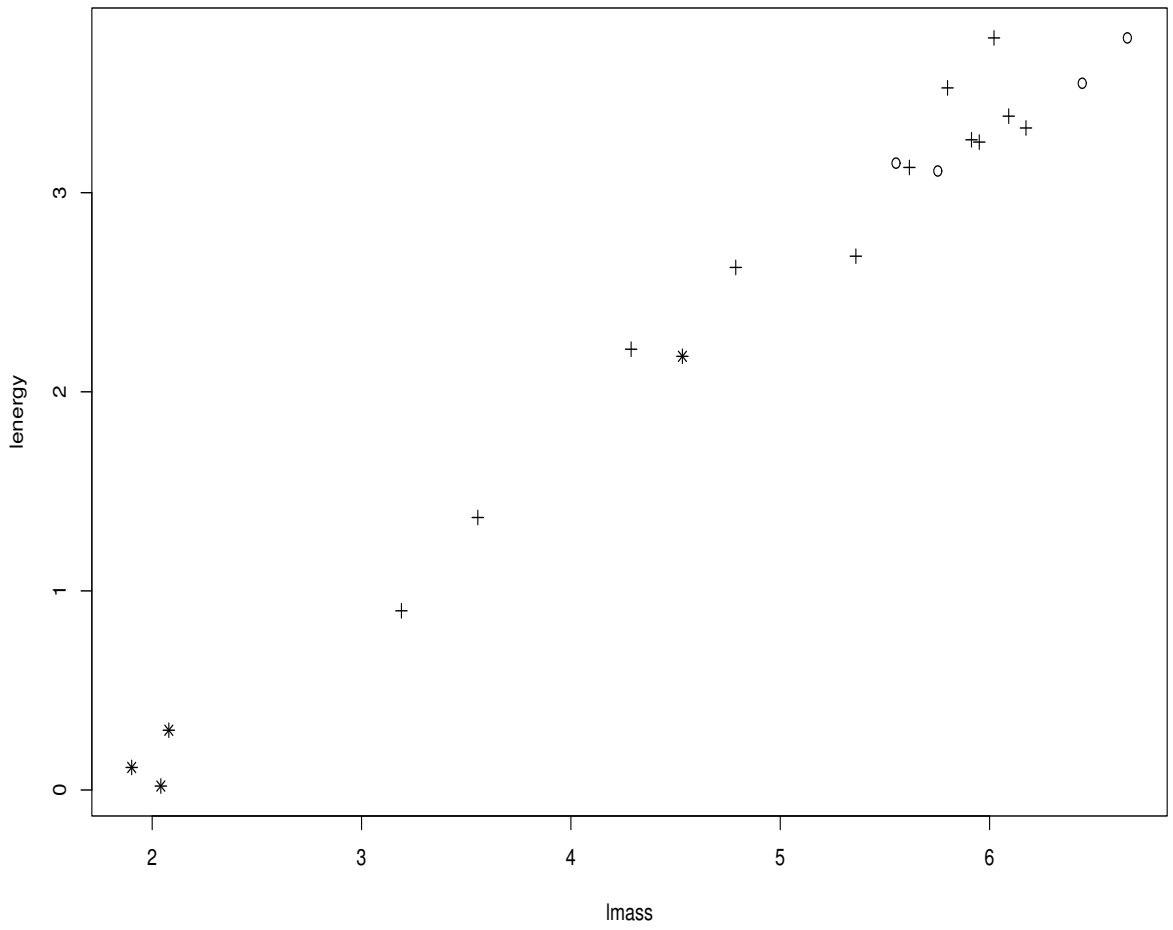

Figure 1: Data

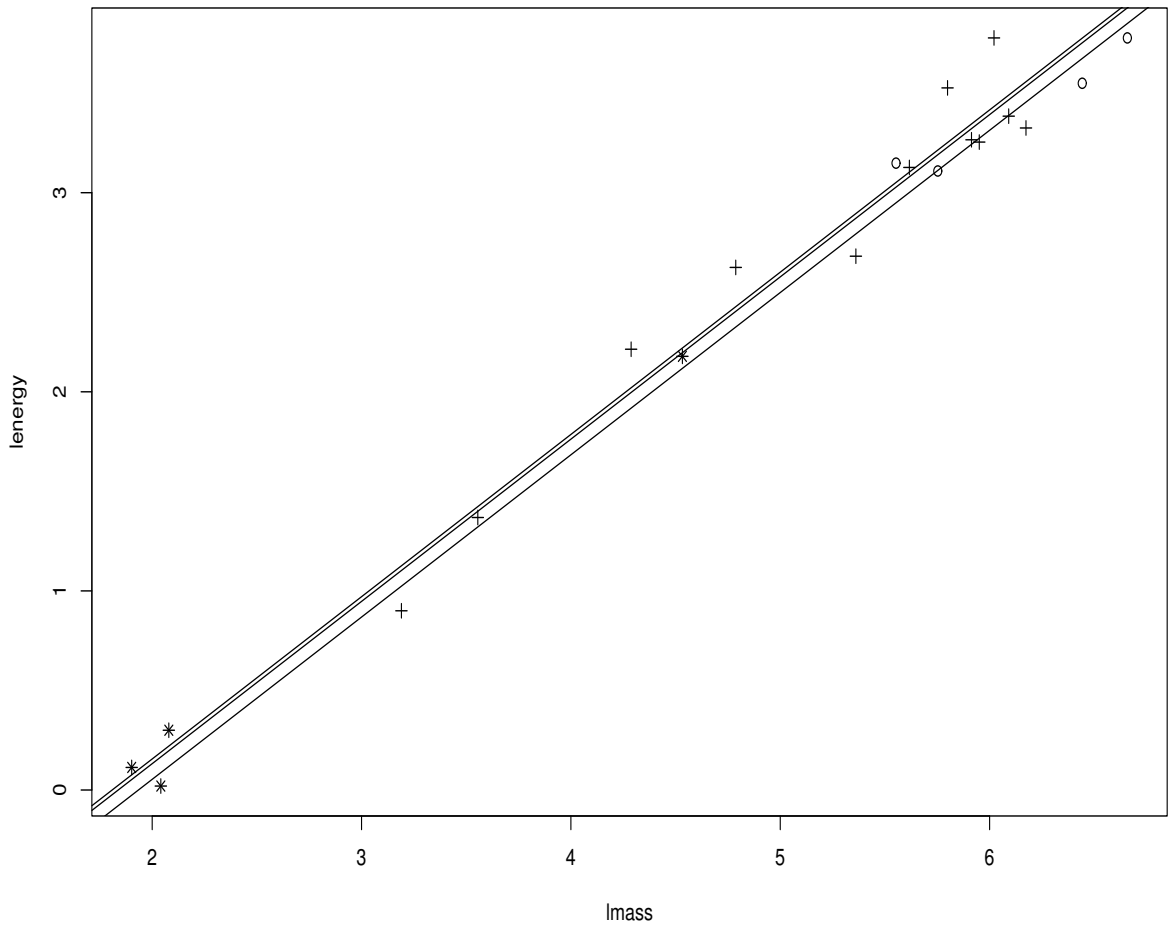

Figure 2: The three regression lines