# Chapter 10 – Inferential Tools for Multiple Regression

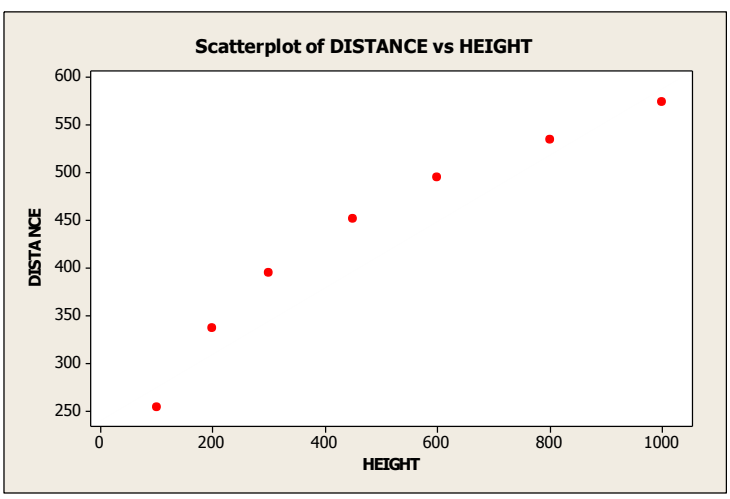
## Case 10.1.1. Galileo's Data on the Motion of Falling Bodies – A Controlled Experiment. *R&S p.272-273.*

**Step 1: Copy the data into a Minitab Worksheet**: use these steps:
File → Open Worksheet →Browse your local directory and upload the csv file Case1001.csv. The data will appear as columns in Minitab with titles DISTANCE and HEIGHT. See ***R&S Display 10.1*** and the data display below:

### Data Display

| Row | DISTANCE | HEIGHT |
|---|---|---|
| 1 | 253 | 100 |
| 2 | 337 | 200 |
| 3 | 395 | 300 |
| 4 | 451 | 450 |
| 5 | 495 | 600 |
| 6 | 534 | 800 |
| 7 | 573 | 1000 |

**Step 2:** Scatterplot of DISTANCE versus HEIGHT is shown below. The scatterplot shows a quadratic relationship between DISTANCE and HEIGHT. This suggests that DISTANCE may be related to both HEIGHT and HEIGHT squared, or HEIGHT$^2$.



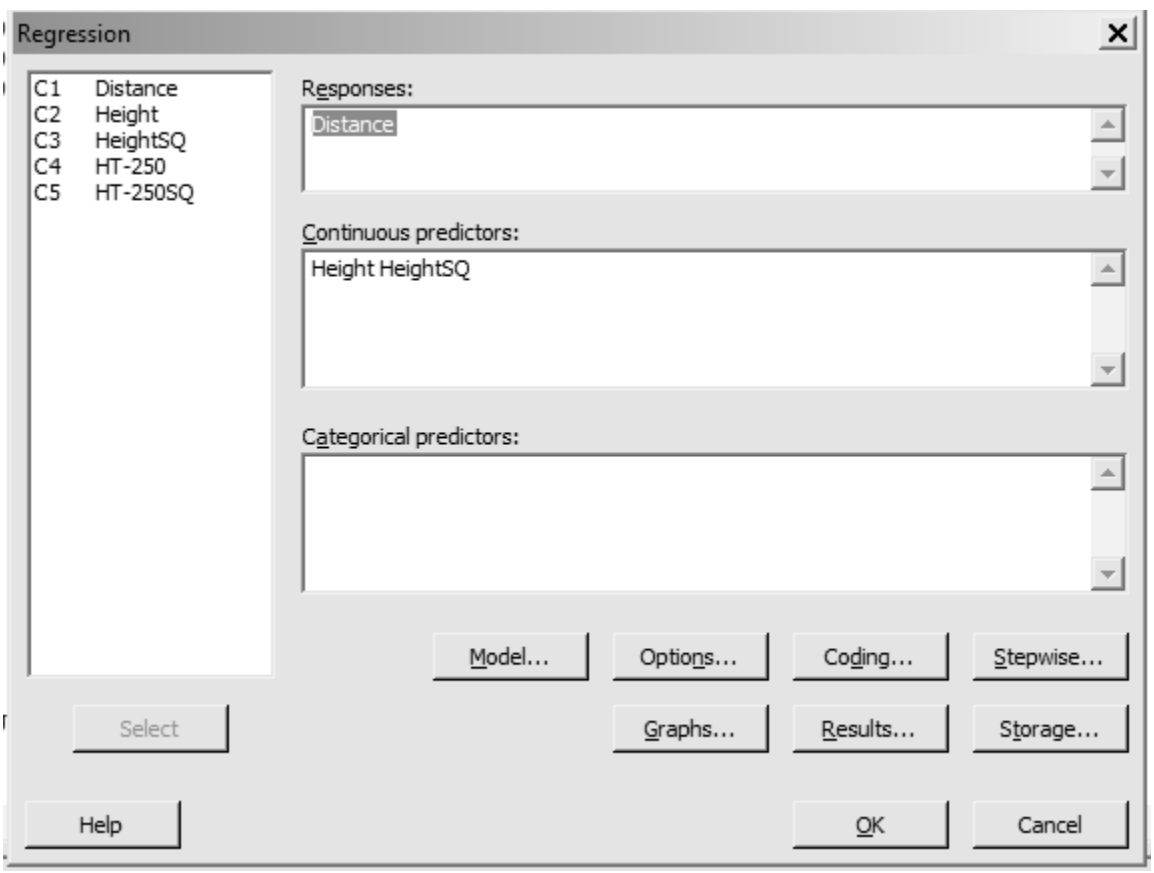

**Step 3:** Create a new variable HEIGHTSQ. Go to Calc and save into column C3 the variable C2**2. Both HEIGHT and HEIGHT$^2$ would be used as predictor variables in a Multiple Regression Model. This is in general called Polynomial Regression, and in this case we may call it a Quadratic Regression Model.

As discussed in R&S p. 279, we may use a reference level of zero for height, in which case we use the predictor variables HEIGHT and HEIGHT$^2$. You may instead choose a reference height of 250 and use the predictor variables HEIGHT – 250 and (HEIGHT-250)$^2$. Using Calc, we compute these new variables and save them in columns C4-C5. The data is displayed below:

**Data Display**

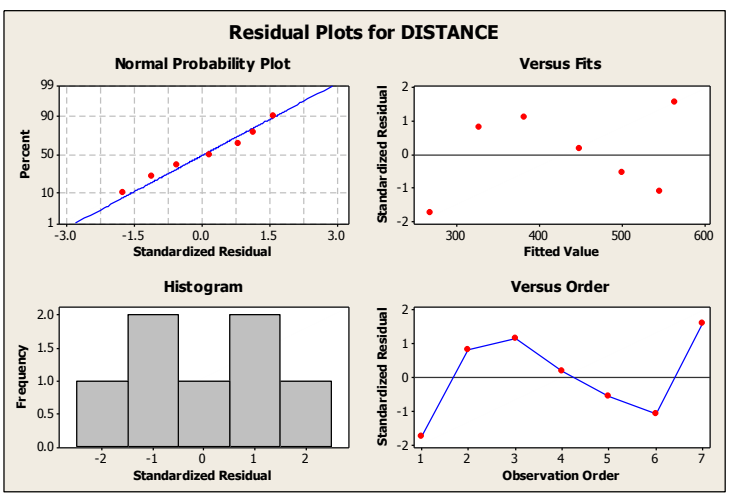
| Row | DISTANCE | HEIGHT | HEIGHTSQ | HT-250 | HT-250SQ |
|---|---|---|---|---|---|
| 1 | 253 | 100 | 10000 | -150 | 22500 |
| 2 | 337 | 200 | 40000 | -50 | 2500 |
| 3 | 395 | 300 | 90000 | 50 | 2500 |
| 4 | 451 | 450 | 202500 | 200 | 40000 |
| 5 | 495 | 600 | 360000 | 350 | 122500 |
| 6 | 534 | 800 | 640000 | 550 | 302500 |
| 7 | 573 | 1000 | 1000000 | 750 | 562500 |

**Step 4:** Fit a Multiple Linear Regression of DISTANCE on two predictor variables, HEIGHT and HEIGHT$^2$. To do this, Go to Stat → Regression; select DISTANCE into Response Variable and select HEIGHT and HEIGHT$^2$ into Predictors.

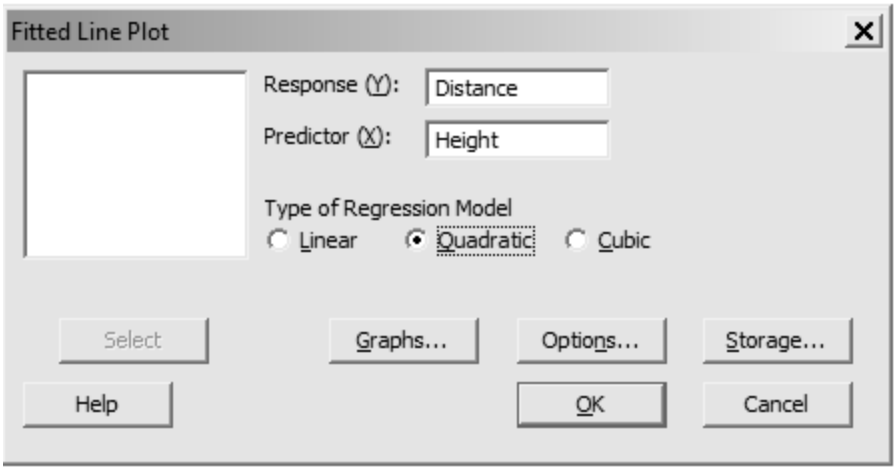


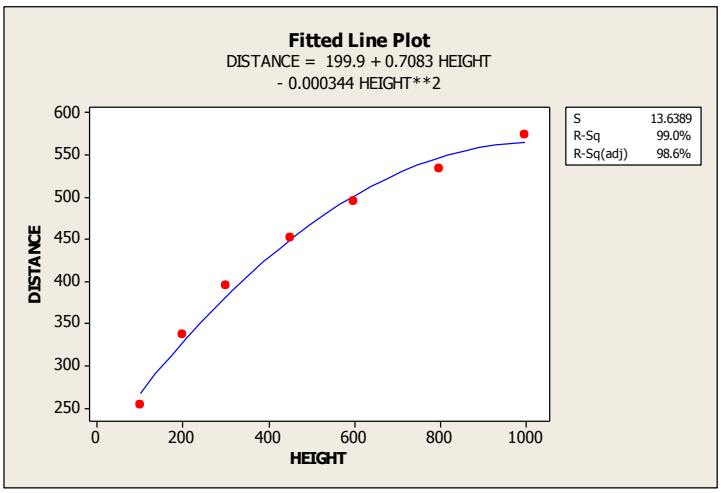

## Regression Analysis: Distance versus Height, HeightSQ

```
Analysis of Variance

Source      DF   Adj SS   Adj MS  F-Value  P-Value
Regression   2  76277.9  38139.0   205.03    0.000
  Height     1  16670.7  16670.7    89.62    0.001
  HeightSQ   1   4927.1   4927.1    26.49    0.007
Error        4    744.1    186.0
Total        6  77022.0


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
13.6389  99.03%     98.55%      92.90%


Coefficients

Term           Coef   SE Coef  T-Value  P-Value    VIF
Constant      199.9      16.8    11.93    0.000
Height       0.7083    0.0748     9.47    0.001  19.33
HeightSQ  -0.000344  0.000067    -5.15    0.007  19.33


Regression Equation

Distance = 199.9 + 0.7083 Height - 0.000344 HeightSQ
```

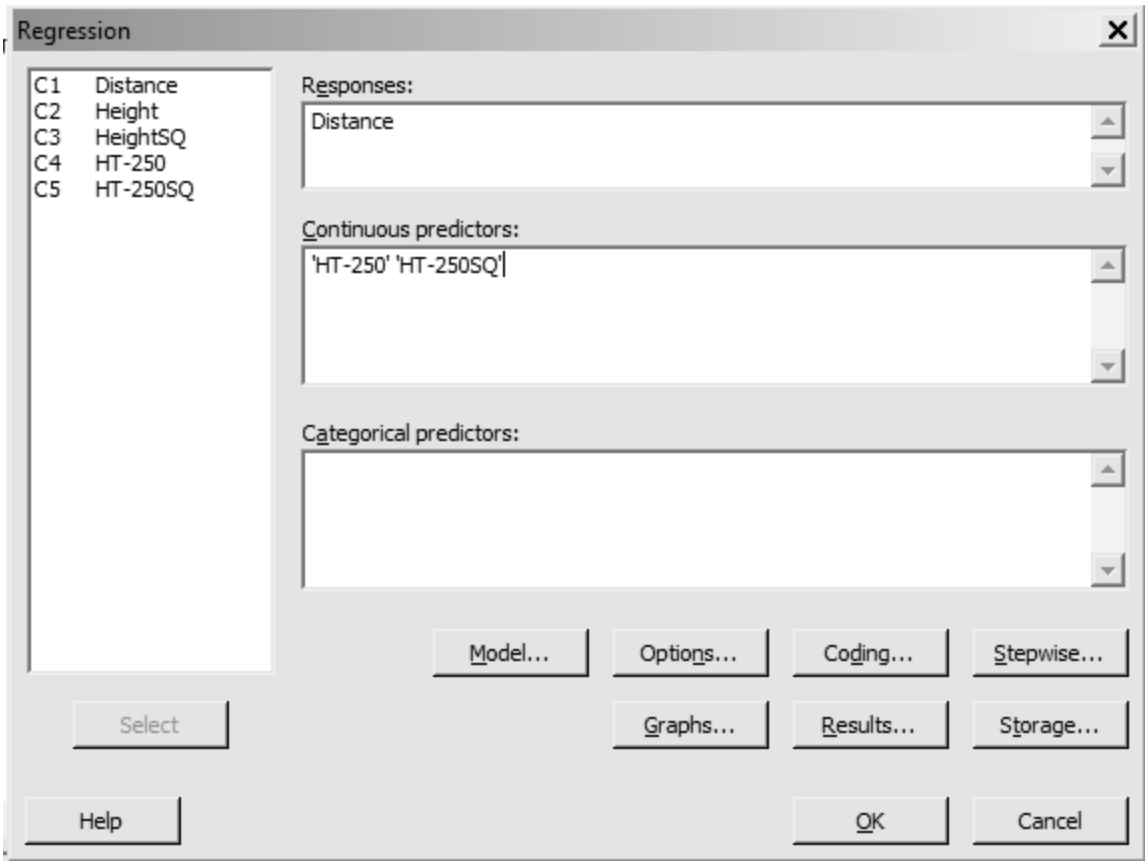



## Fitted line plot



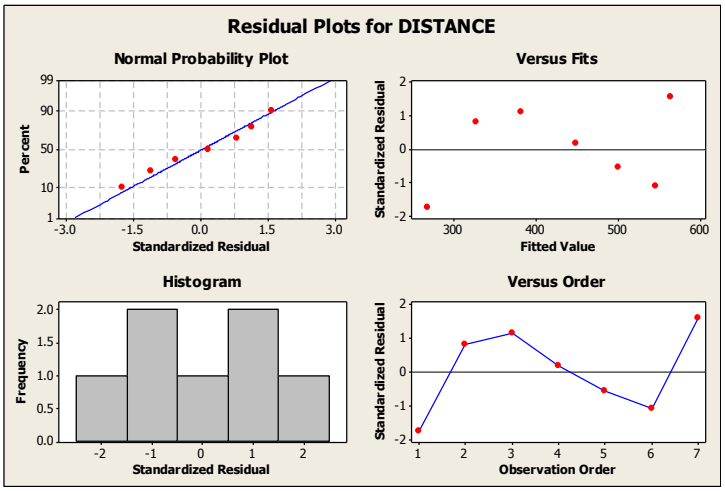

**Step 5:** Fit a Multiple Linear Regression of DISTANCE on two predictor variables, HT-250 and HT-250SQ. To do this, Go to Stat → Regression; select DISTANCE into Response Variable and select HT-250 and HT-250SQ into Predictors.

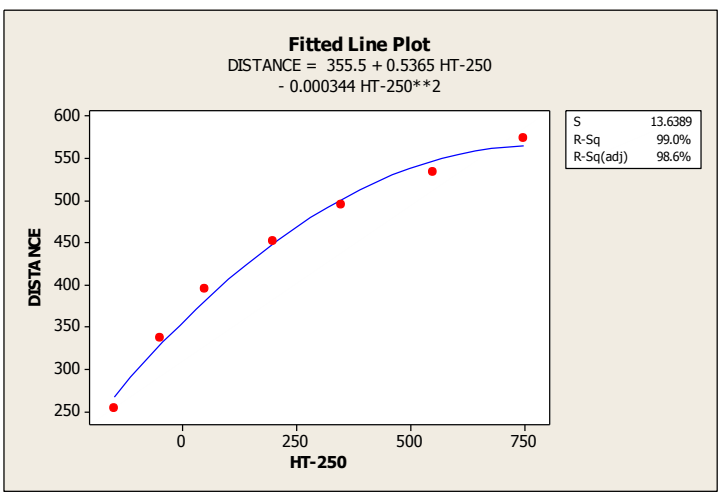



## Regression Analysis: Distance versus HT-250, HT-250SQ

```
Analysis of Variance

Source       DF   Adj SS   Adj MS  F-Value  P-Value
Regression    2  76277.9  38139.0   205.03    0.000
  HT-250      1  28976.5  28976.5   155.77    0.000
  HT-250SQ    1   4927.1   4927.1    26.49    0.007
Error         4    744.1    186.0
Total         6  77022.0


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
13.6389  99.03%     98.55%      92.90%


Coefficients

Term           Coef   SE Coef  T-Value  P-Value   VIF
Constant     355.51      6.62    53.66    0.000
HT-250       0.5365    0.0430    12.48    0.000  6.38
HT-250SQ  -0.000344  0.000067    -5.15    0.007  6.38


Regression Equation
```

```
Distance = 355.51 + 0.5365 HT-250 - 0.000344 HT-250SQ
```

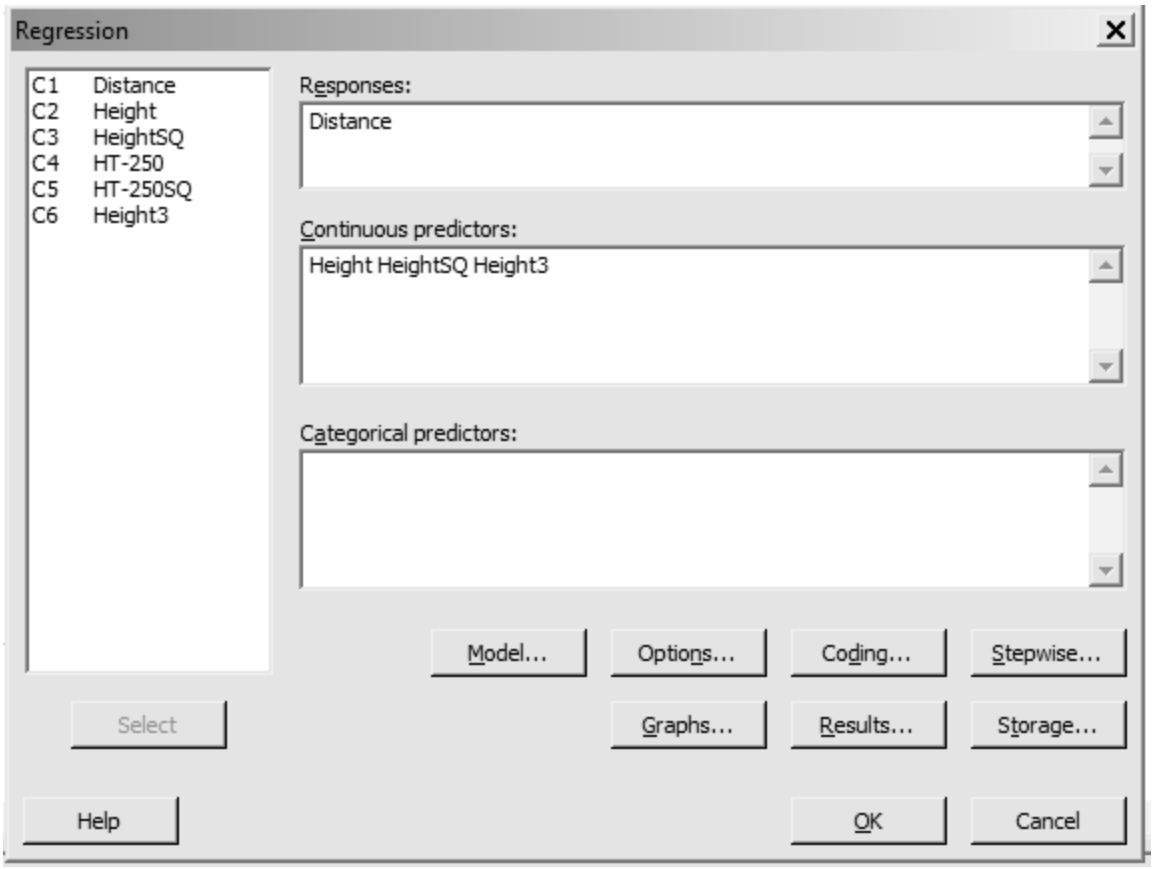



Is there a reason to include a cubic term in Height? Investigate using an Extra-SS-F-test and change in $R^2$ statistic:

$$\mu\{\text{DISTANCE} \mid \text{HEIGHT}, \text{HEIGHT}^2, \text{HEIGHT}^3\} = \beta_0 + \beta_1\ \text{HEIGHT} + \beta_2\ \text{HEIGHT}^2 + \beta_3\ \text{HEIGHT}^3$$

**Step 6:** Create a new variable HEIGHT3. Go to Calc and save into column C10 the variable C2**3. The three variables HEIGHT, $\text{HEIGHT}^2$ and $\text{HEIGHT}^3$ are included as predictors in the multiple regression model.

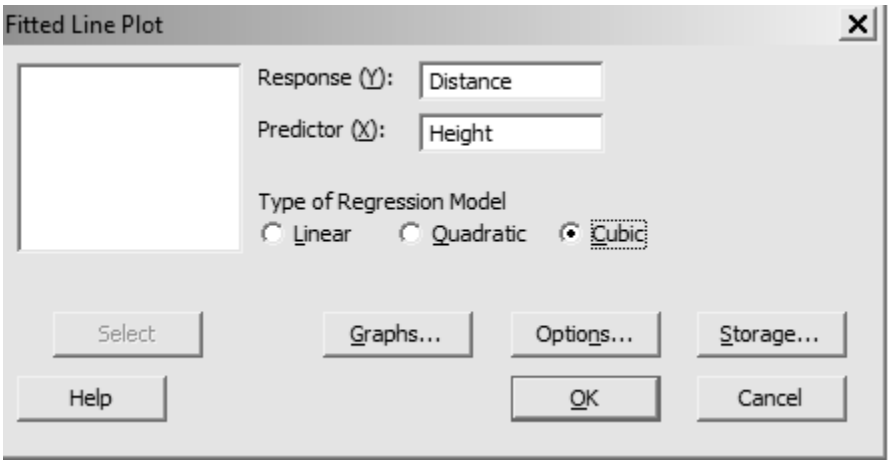


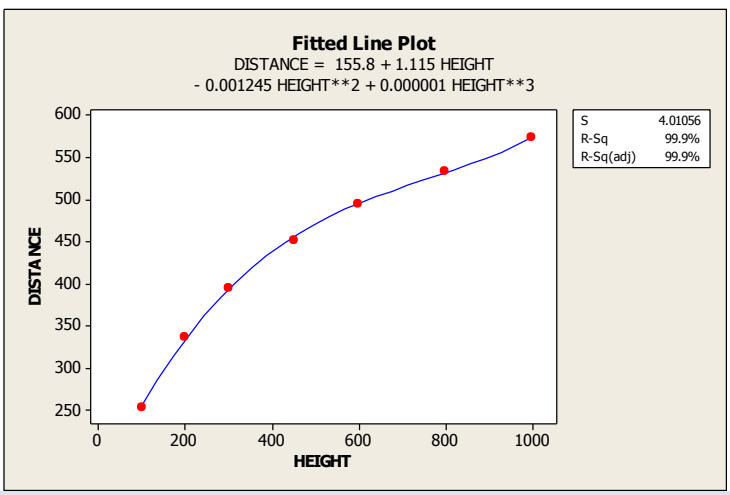

## Regression Analysis: Distance versus Height, HeightSQ, Height3

```
Analysis of Variance

Source      DF   Adj SS   Adj MS  F-Value  P-Value
Regression   3  76973.7  25657.9  1595.19    0.000
  Height     1   4639.1   4639.1   288.42    0.000
  HeightSQ   1   1301.0   1301.0    80.89    0.003
  Height3    1    695.8    695.8    43.26    0.007
Error        3     48.3     16.1
Total        6  77022.0


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
4.01056  99.94%     99.87%      98.58%


Coefficients

Term           Coef   SE Coef  T-Value  P-Value     VIF
Constant     155.78      8.33    18.71    0.000
Height       1.1153    0.0657    16.98    0.000  172.18
HeightSQ  -0.001245  0.000138    -8.99    0.003  960.31
Height3    0.000001  0.000000     6.58    0.007  356.22


Regression Equation
```

```
Distance = 155.78 + 1.1153 Height - 0.001245 HeightSQ + 0.000001 Height3
```

**Fitted line Plot – Cubic Fit**



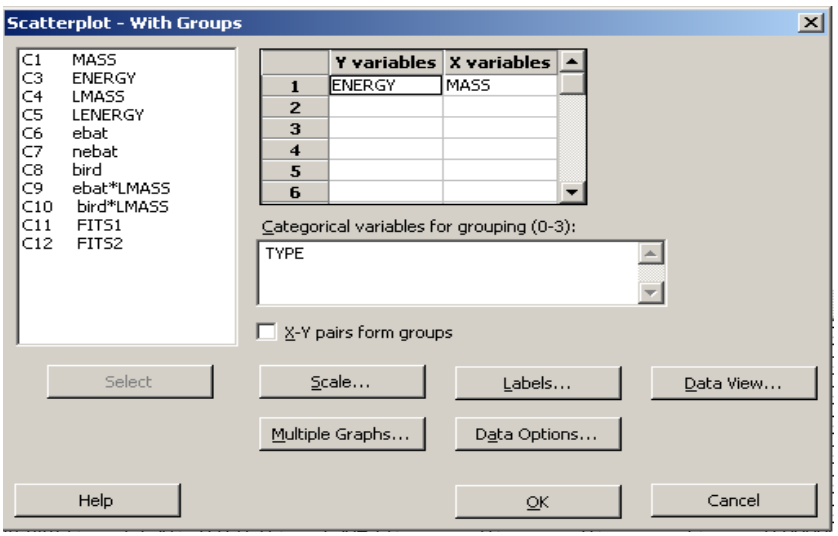

## Case 10.1.2. The Energy Costs of Echolocation by Bats – An Observational Study. *R&S p.273-275.*

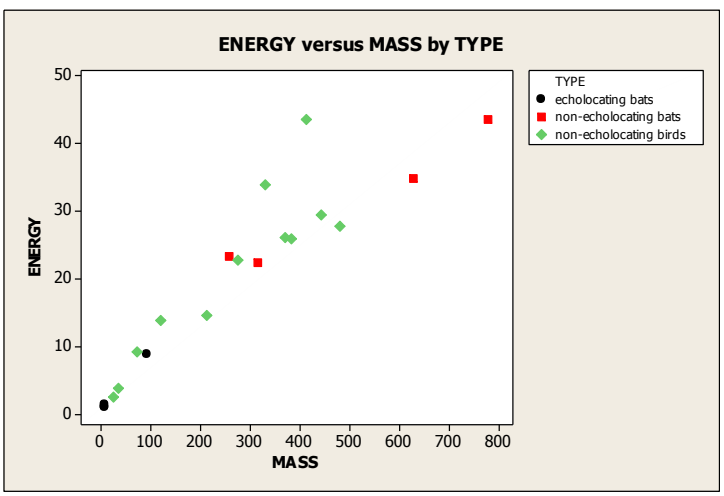
**Step 1: Copy the data into a Minitab Worksheet**: use these steps: File → Open Worksheet →Browse your local directory and upload the csv file Case1002.csv. To display the data in Minitab, go to Data →Display Data, and copy the columns C1-C5 in the window on the right. The data will appear as columns in Minitab. ***See R&S Display 10.3*** for data display on 20 different species. Do a scatterplot of ENERGY versus MASS by TYPE.

### Data Display

| Row | MASS | TYPE | ENERGY |
|---|---|---|---|
| 1 | 779.000 | non-echolocating bats | 43.7000 |
| 2 | 628.000 | non-echolocating bats | 34.8000 |
| 3 | 258.000 | non-echolocating bats | 23.3000 |
| 4 | 315.000 | non-echolocating bats | 22.4000 |
| 5 | 24.300 | non-echolocating birds | 2.4600 |
| 6 | 35.000 | non-echolocating birds | 3.9300 |

```
  7   72.800  non-echolocating birds   9.1500
  8  120.000  non-echolocating birds  13.8000
  9  213.000  non-echolocating birds  14.6000
 10  275.000  non-echolocating birds  22.8000
 11  370.000  non-echolocating birds  26.2000
 12  384.000  non-echolocating birds  25.9000
 13  442.000  non-echolocating birds  29.5000
 14  412.000  non-echolocating birds  43.7000
 15  330.000  non-echolocating birds  34.0000
 16  480.000  non-echolocating birds  27.8000
 17   93.000  echolocating bats        8.8300
 18    8.000  echolocating bats        1.3500
 19    6.700  echolocating bats        1.1200
 20    7.700  echolocating bats        1.0200
```



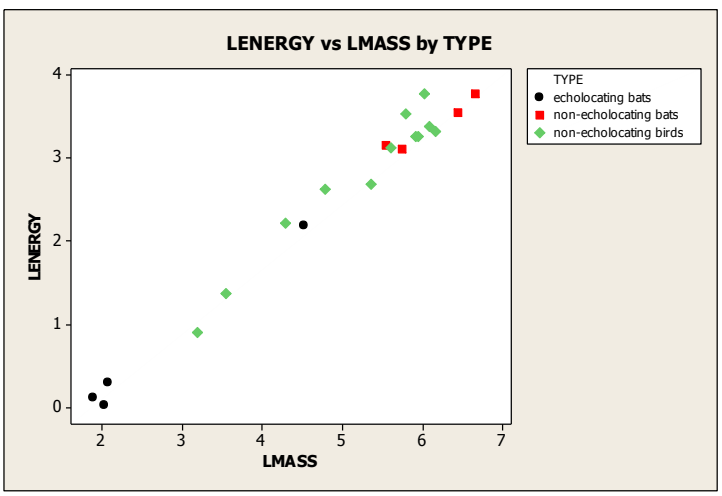

**Step 2:** Construct log-transformed variables LENERGY and LMASS, and display all data.

## Data Display

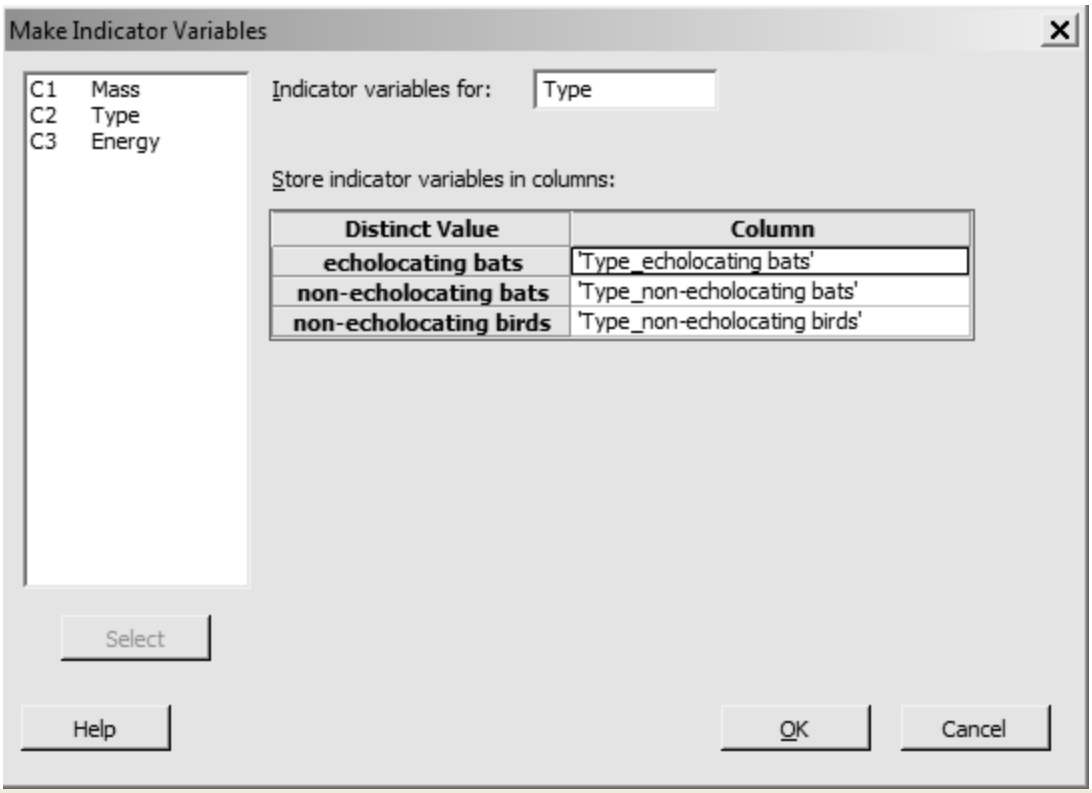
```
Row     MASS  TYPE                      ENERGY    LMASS  LENERGY
  1  779.000  non-echolocating bats    43.7000  6.65801  3.77735
  2  628.000  non-echolocating bats    34.8000  6.44254  3.54962
```

| | | | | | |
|---|---|---|---|---|---|
| 3 | 258.000 | non-echolocating bats | 23.3000 | 5.55296 | 3.14845 |
| 4 | 315.000 | non-echolocating bats | 22.4000 | 5.75257 | 3.10906 |
| 5 | 24.300 | non-echolocating birds | 2.4600 | 3.19048 | 0.90016 |
| 6 | 35.000 | non-echolocating birds | 3.9300 | 3.55535 | 1.36864 |
| 7 | 72.800 | non-echolocating birds | 9.1500 | 4.28772 | 2.21375 |
| 8 | 120.000 | non-echolocating birds | 13.8000 | 4.78749 | 2.62467 |
| 9 | 213.000 | non-echolocating birds | 14.6000 | 5.36129 | 2.68102 |
| 10 | 275.000 | non-echolocating birds | 22.8000 | 5.61677 | 3.12676 |
| 11 | 370.000 | non-echolocating birds | 26.2000 | 5.91350 | 3.26576 |
| 12 | 384.000 | non-echolocating birds | 25.9000 | 5.95064 | 3.25424 |
| 13 | 442.000 | non-echolocating birds | 29.5000 | 6.09131 | 3.38439 |
| 14 | 412.000 | non-echolocating birds | 43.7000 | 6.02102 | 3.77735 |
| 15 | 330.000 | non-echolocating birds | 34.0000 | 5.79909 | 3.52636 |
| 16 | 480.000 | non-echolocating birds | 27.8000 | 6.17379 | 3.32504 |
| 17 | 93.000 | echolocating bats | 8.8300 | 4.53260 | 2.17816 |
| 18 | 8.000 | echolocating bats | 1.3500 | 2.07944 | 0.30010 |
| 19 | 6.700 | echolocating bats | 1.1200 | 1.90211 | 0.11333 |
| 20 | 7.700 | echolocating bats | 1.0200 | 2.04122 | 0.01980 |

**Step 3:** Scatterplot of LENERGY versus LMASS by TYPE. Go to Graph→ Scatterplots → with groups. Also see ***R&S Display 10.4.***




**Step 4:** Create Indicator Variables. Create Indicator variables corresponding to the categorical variable TYPE. Go to Calc → Make Indicator Variables, and select TYPE. This creates three new columns C6-C8, which we name ebat, nebat, and bird. Note that only two of the three indicator variables can be used as predictors.



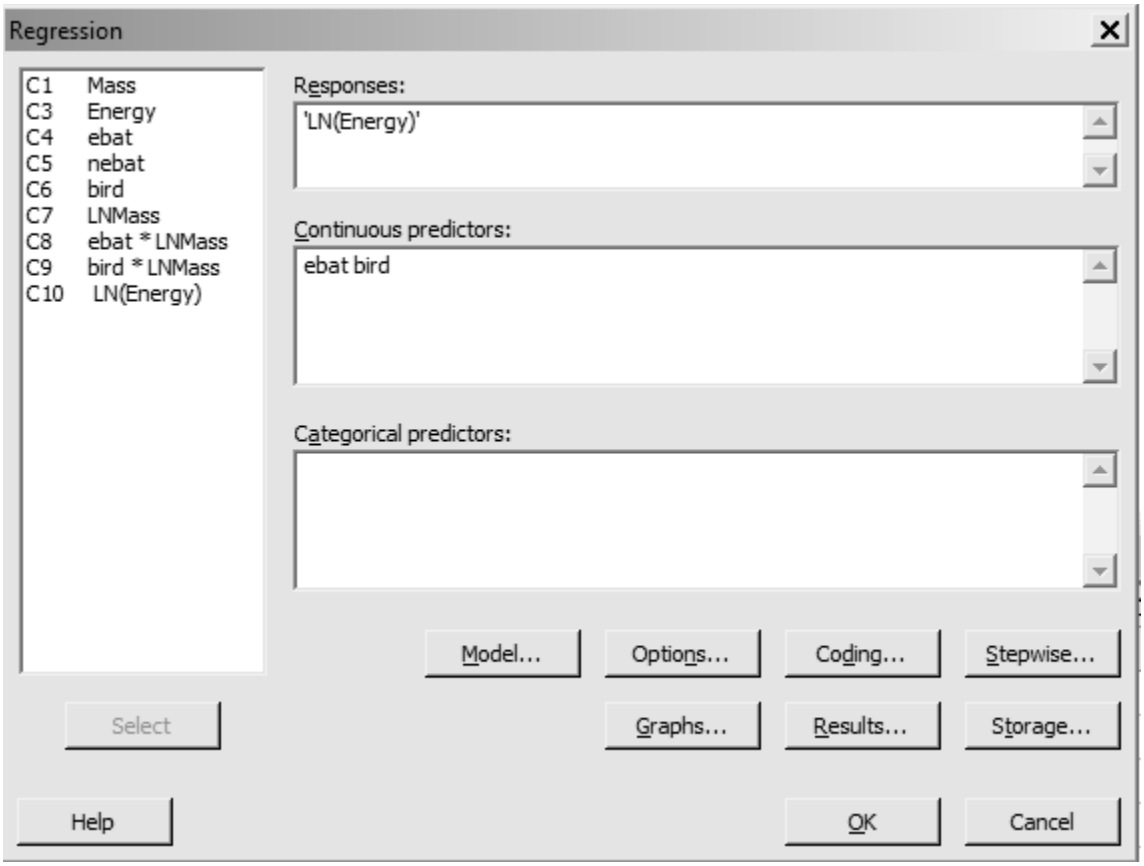

**Data Display**

| Row | TYPE | ebat | nebat | bird |
|---|---|---|---|---|
| 1 | non-echolocating bats | 0 | 1 | 0 |
| 2 | non-echolocating bats | 0 | 1 | 0 |
| 3 | non-echolocating bats | 0 | 1 | 0 |
| 4 | non-echolocating bats | 0 | 1 | 0 |
| 5 | non-echolocating birds | 0 | 0 | 1 |
| 6 | non-echolocating birds | 0 | 0 | 1 |
| 7 | non-echolocating birds | 0 | 0 | 1 |
| 8 | non-echolocating birds | 0 | 0 | 1 |
| 9 | non-echolocating birds | 0 | 0 | 1 |
| 10 | non-echolocating birds | 0 | 0 | 1 |
| 11 | non-echolocating birds | 0 | 0 | 1 |
| 12 | non-echolocating birds | 0 | 0 | 1 |
| 13 | non-echolocating birds | 0 | 0 | 1 |
| 14 | non-echolocating birds | 0 | 0 | 1 |
| 15 | non-echolocating birds | 0 | 0 | 1 |
| 16 | non-echolocating birds | 0 | 0 | 1 |
| 17 | echolocating bats | 1 | 0 | 0 |
| 18 | echolocating bats | 1 | 0 | 0 |
| 19 | echolocating bats | 1 | 0 | 0 |
| 20 | echolocating bats | 1 | 0 | 0 |

**Step 5:** Create Interaction Variables ebat*LMASS and bird*LMASS, by going to Calc. Save the Interaction variables in columns C9 and C10.

**Step 6: Multiple Regression Modeling. As discussed in *R&S sec. 10.2*,** we fit three different Multiple Regression Models.

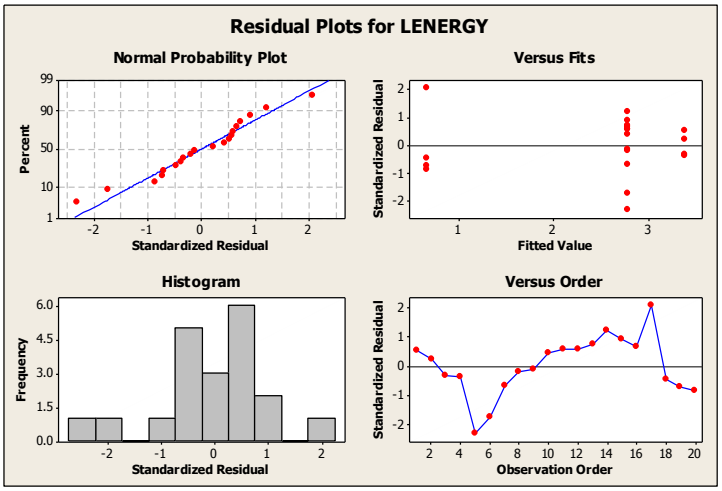
1. ***Model (1).*** Go to Stat → Regression; select LENERGY into Response window and select ebat and bird into the Predictors window; select Residuals Plots for graphs and click ok.



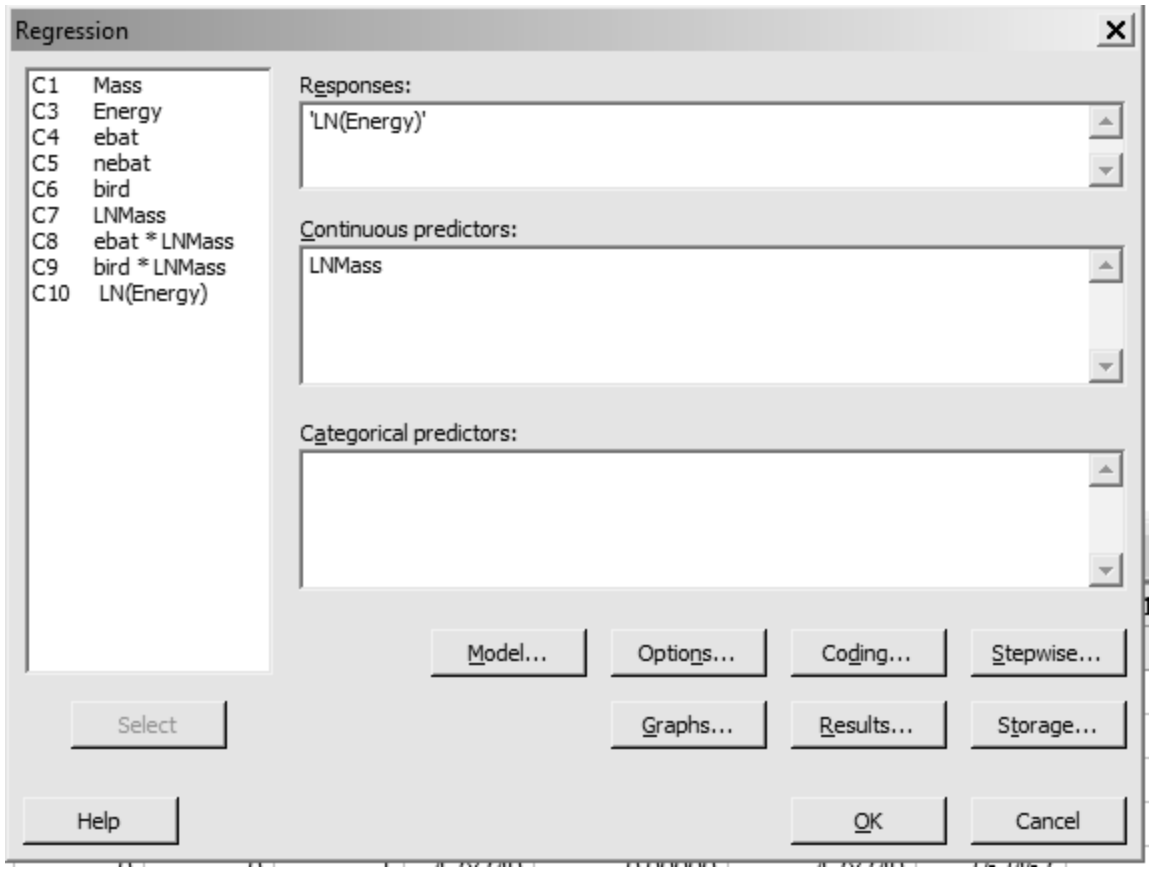

## Regression Analysis: LNENERGY versus ebat, bird

```
Analysis of Variance

Source      DF  Adj SS   Adj MS  F-Value  P-Value
Regression   2  17.844   8.9222    12.50    0.000
  ebat       1  15.051  15.0511    21.09    0.000
  bird       1   1.112   1.1118     1.56    0.229
Error       17  12.130   0.7135
Total       19  29.975


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.844718  59.53%     54.77%      45.06%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   3.396    0.422     8.04    0.000
```

```
ebat       -2.743    0.597    -4.59    0.000  1.60
bird       -0.609    0.488    -1.25    0.229  1.60


Regression Equation

LNENERGY = 3.396 - 2.743 ebat - 0.609 bird


Fits and Diagnostics for Unusual Observations

Obs  LNENERGY    Fit   Resid  Std Resid
  5     0.900  2.787  -1.887      -2.33  R
 17     2.178  0.653   1.525       2.09  R

R  Large residual
```



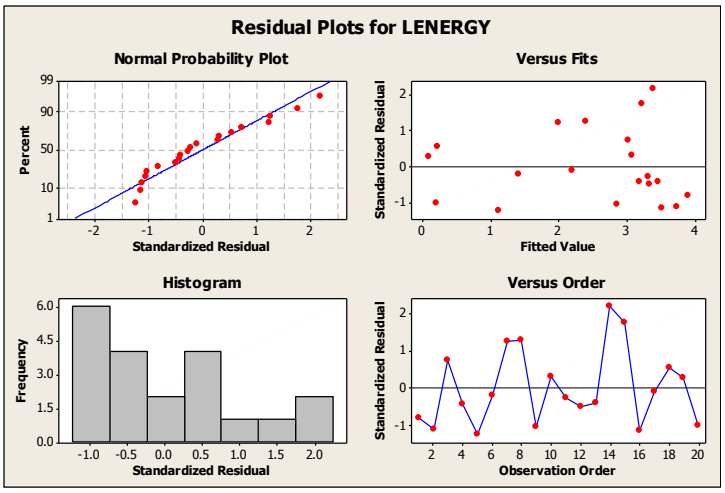

2. ***Model (2).*** Go to Stat → Regression; select LENERGY into Response window and select LMASS into the Predictors window; select Residuals Plots for graphs and click ok.



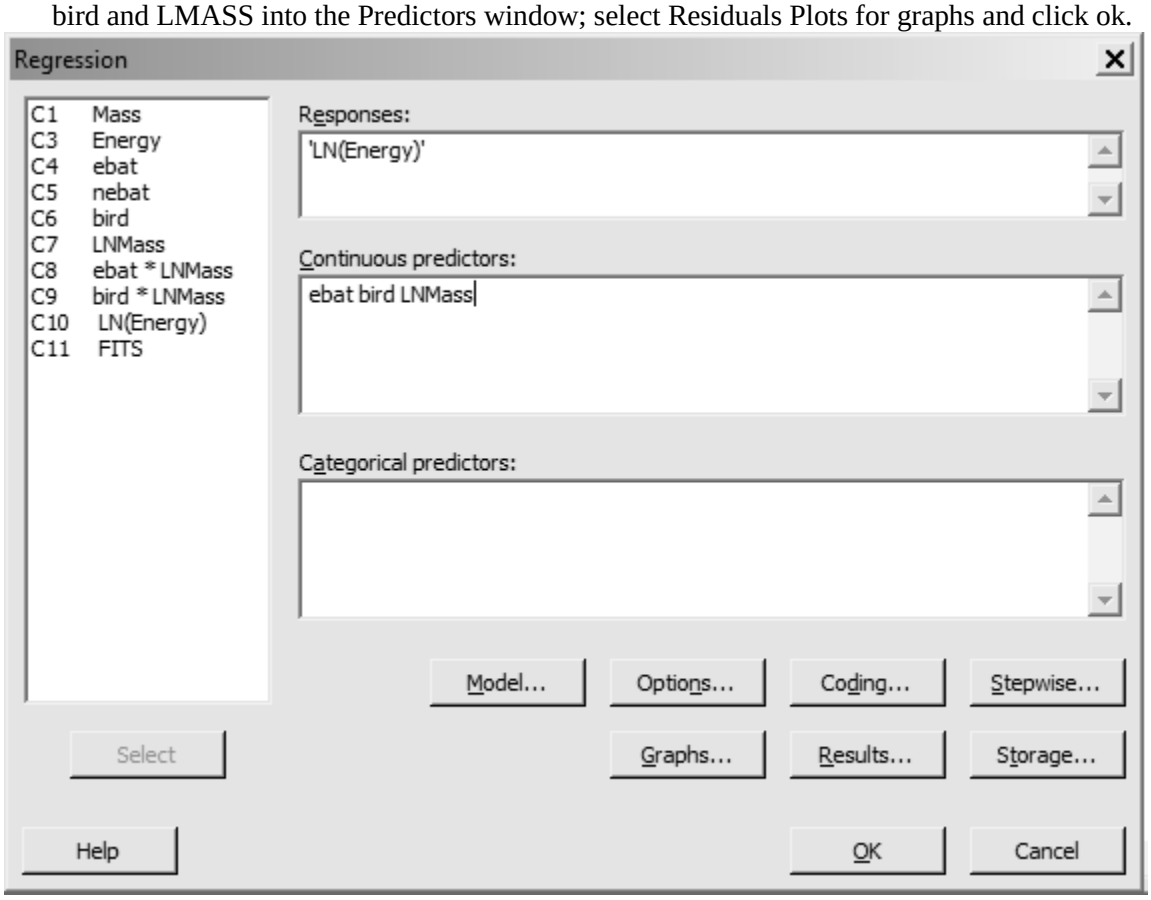

## Regression Analysis: LN(Energy) versus LNMass

```
Analysis of Variance

Source      DF   Adj SS   Adj MS  F-Value  P-Value
Regression   1  29.3919  29.3919   907.64    0.000
  LNMass     1  29.3919  29.3919   907.64    0.000
Error       18   0.5829   0.0324
Total       19  29.9748


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.179952  98.06%     97.95%      97.65%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant  -1.468    0.137   -10.70    0.000
LNMass    0.8086   0.0268    30.13    0.000  1.00


Regression Equation

LN(Energy) = -1.468 + 0.8086 LNMass


Fits and Diagnostics for Unusual Observations
```

```
                                    Std
Obs   LN(Energy)      Fit    Resid   Resid
 14       3.7773   3.4004   0.3769    2.18   R

R  Large residual
```

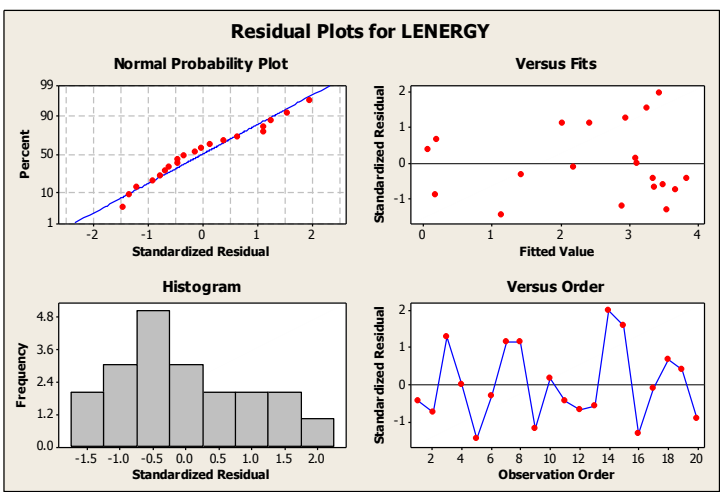



3. ***Model (3).*** Go to Stat → Regression; select LENERGY into Response window and select ebat, bird and LMASS into the Predictors window; select Residuals Plots for graphs and click ok.

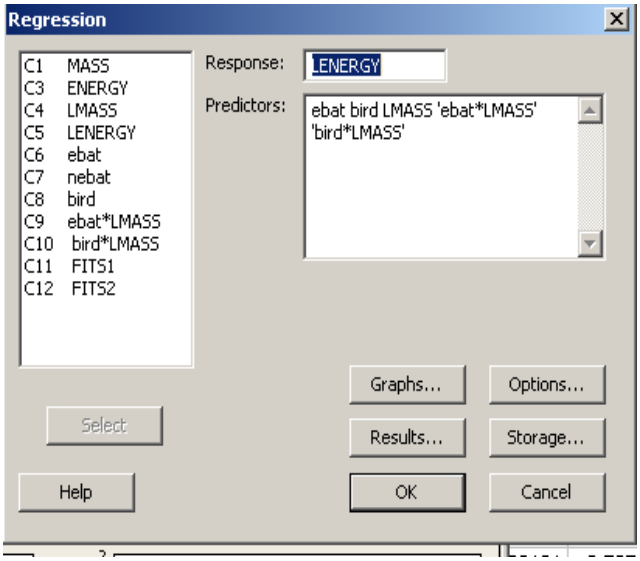


## Regression Analysis: LNENERGY versus ebat, bird, LNMASS

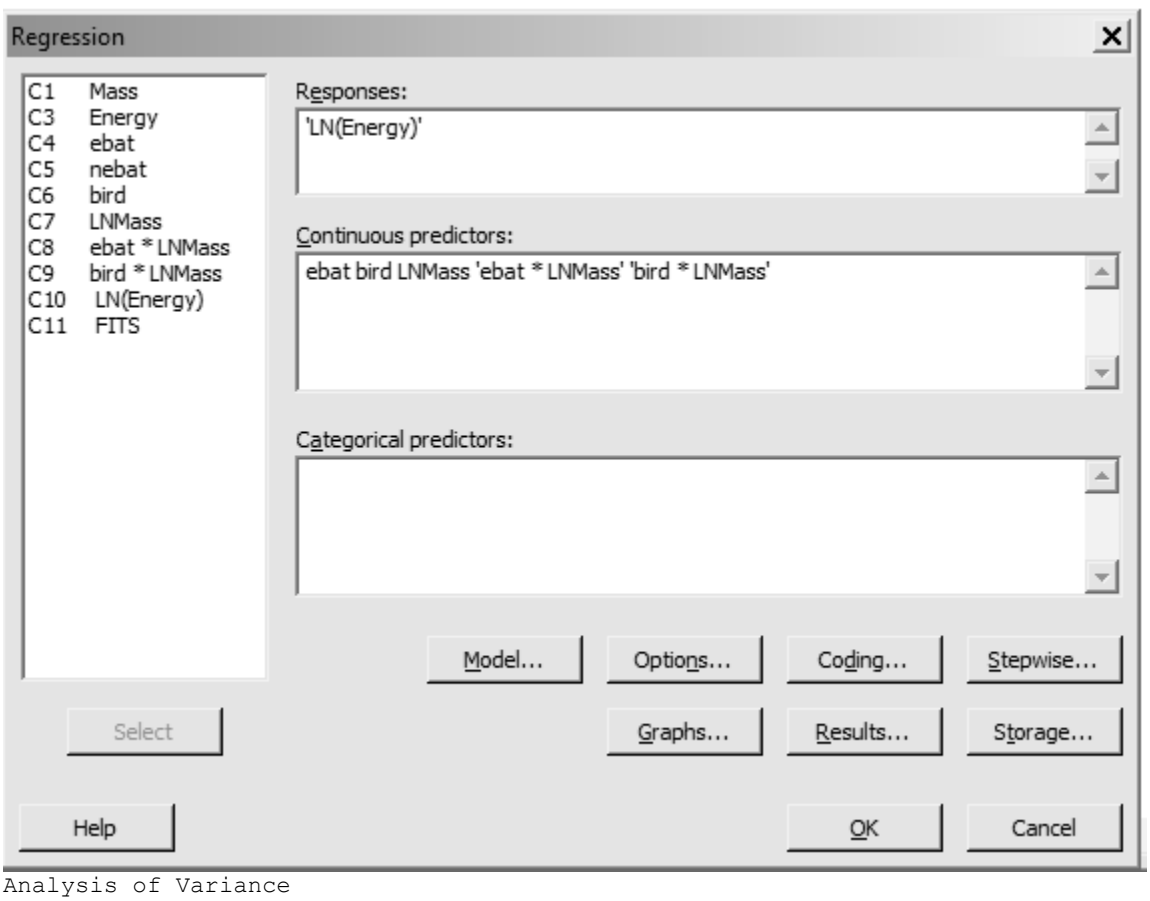
```
Analysis of Variance

Source       DF   Adj SS   Adj MS  F-Value  P-Value
Regression    3  29.4215   9.8072   283.59    0.000
  ebat        1   0.0052   0.0052     0.15    0.703
  bird        1   0.0277   0.0277     0.80    0.384
  LNMASS      1  11.5770  11.5770   334.77    0.000
Error        16   0.5533   0.0346
Total        19  29.9748


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.185963  98.15%     97.81%      97.30%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant  -1.576    0.287    -5.49    0.000
ebat       0.079    0.203     0.39    0.703  3.80
bird       0.102    0.114     0.90    0.384  1.81
LNMASS    0.8150   0.0445    18.30    0.000  2.58


Regression Equation

LNENERGY = -1.576 + 0.079 ebat + 0.102 bird + 0.8150 LNMASS
```




4. ***Model (3).*** Go to Stat → Regression; select LENERGY into Response window and select ebat, bird, LMASS, ebat*LMASS and bird*LMASS into the Predictors window; select Residuals Plots for graphs and click ok.




**Regression Analysis: LNENERGY versus ebat, bird, LNMASS, ebat * LMASS, bird * LMASS**




```
Analysis of Variance

Source          DF   Adj SS   Adj MS  F-Value  P-Value
Regression       5  29.4699  5.89399   163.44    0.000
  ebat           1   0.0351  0.03509     0.97    0.341
```

| | | | | | |
|---|---|---|---|---|---|
| bird | 1 | 0.0408 | 0.04084 | 1.13 | 0.305 |
| LNMASS | 1 | 0.2952 | 0.29520 | 8.19 | 0.013 |
| ebat * LMASS | 1 | 0.0333 | 0.03330 | 0.92 | 0.353 |
| bird * LMASS | 1 | 0.0477 | 0.04775 | 1.32 | 0.269 |
| Error | 14 | 0.5049 | 0.03606 | | |
| Total | 19 | 29.9748 | | | |

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 0.189900 | 98.32% | 97.71% | 96.29% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | -0.20 | 1.26 | -0.16 | 0.875 | |
| ebat | -1.27 | 1.29 | -0.99 | 0.341 | 146.62 |
| bird | -1.38 | 1.30 | -1.06 | 0.305 | 223.30 |
| LNMASS | 0.590 | 0.206 | 2.86 | 0.013 | 52.97 |
| ebat * LMASS | 0.215 | 0.224 | 0.96 | 0.353 | 37.56 |
| bird * LMASS | 0.246 | 0.213 | 1.15 | 0.269 | 180.67 |

Regression Equation

LNENERGY = -0.20 - 1.27 ebat - 1.38 bird + 0.590 LNMASS + 0.215 ebat * LMASS
+ 0.246 bird * LMASS

Fits and Diagnostics for Unusual Observations

| Obs | LNENERGY | Fit | Resid | Std Resid | |
|---|---|---|---|---|---|
| 17 | 2.178 | 2.177 | 0.001 | 0.15 | X |

X Unusual X