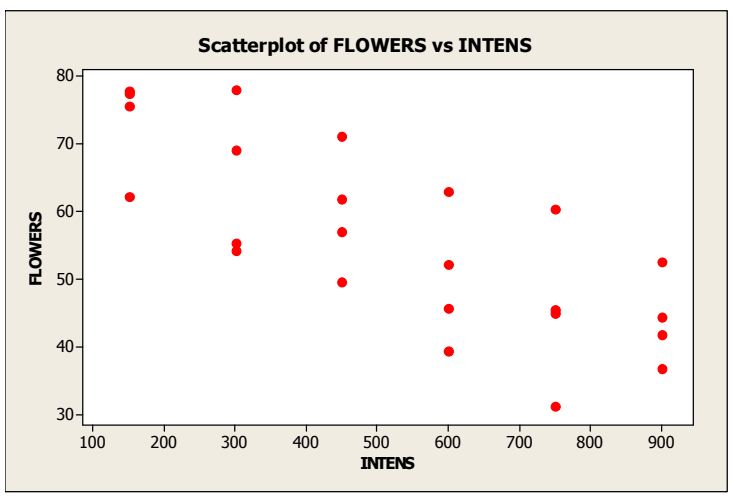
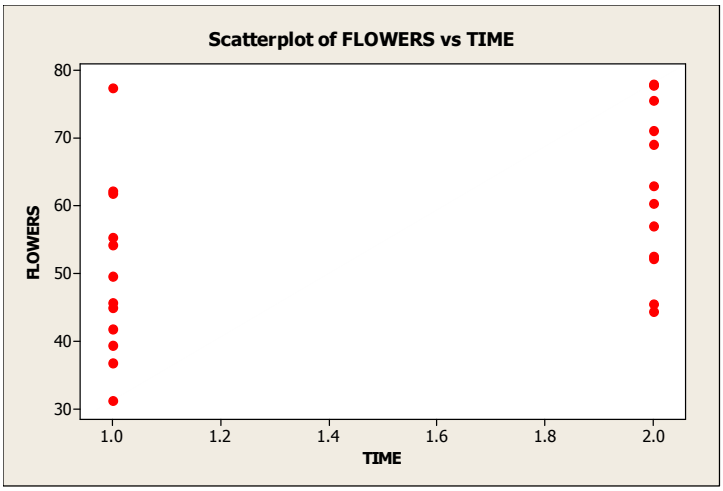
# Chapter 9 – Multiple Regression

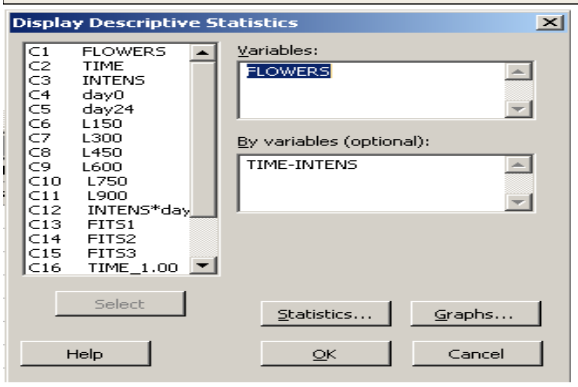
## Case 9.1.1. Effects of Light on Meadowfoam Flowering – A Randomized Experiment *R&S p.238-239.*

**Step 1: Copy the data into a Minitab Worksheet**: use these steps:
File → Open Worksheet →Browse your local directory and upload the csv file case0901.csv. The data will appear as columns in Minitab with titles FLOWERS, TIME, and INTENS. Note that both TIME and INTENS are categorical predictor variables.
The categorical predictor TIME has 2 levels: level 1 indicates timing level or day 0, i.e., Late (at PFI), and level 2 indicates timing level or day 24, i.e., Early (24 days before PFI).
The categorical predictor INTENS denotes light intensity ($\mu$mol/m$^2$/sec) at 6 levels: 150, 300, 450, 600, 750, and 900. ***See R&S Display 9.1.*** The data is shown in tabular form in ***R&S Display 9.2.***

**Step 2:** Scatterplots of FLOWER versus INTENS and FLOWER versus TIME are shown below, as well as summary statistics for FLOWER by TIME and INTENS.






## Descriptive Statistics: FLOWERS

### Results for TIME = 1.00

| Variable | INTENS | N | N* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 |
|---|---|---|---|---|---|---|---|---|---|---|
| FLOWERS | 150.00 | 2 | 0 | 69.85 | 7.55 | 10.68 | 62.30 | * | 69.85 | * |
| | 300.00 | 2 | 0 | 54.750 | 0.550 | 0.778 | 54.200 | * | 54.750 | * |
| | 450.00 | 2 | 0 | 55.75 | 6.15 | 8.70 | 49.60 | * | 55.75 | * |
| | 600.00 | 2 | 0 | 42.55 | 3.15 | 4.45 | 39.40 | * | 42.55 | * |
| | 900.00 | 2 | 0 | 39.35 | 2.55 | 3.61 | 36.80 | * | 39.35 | * |

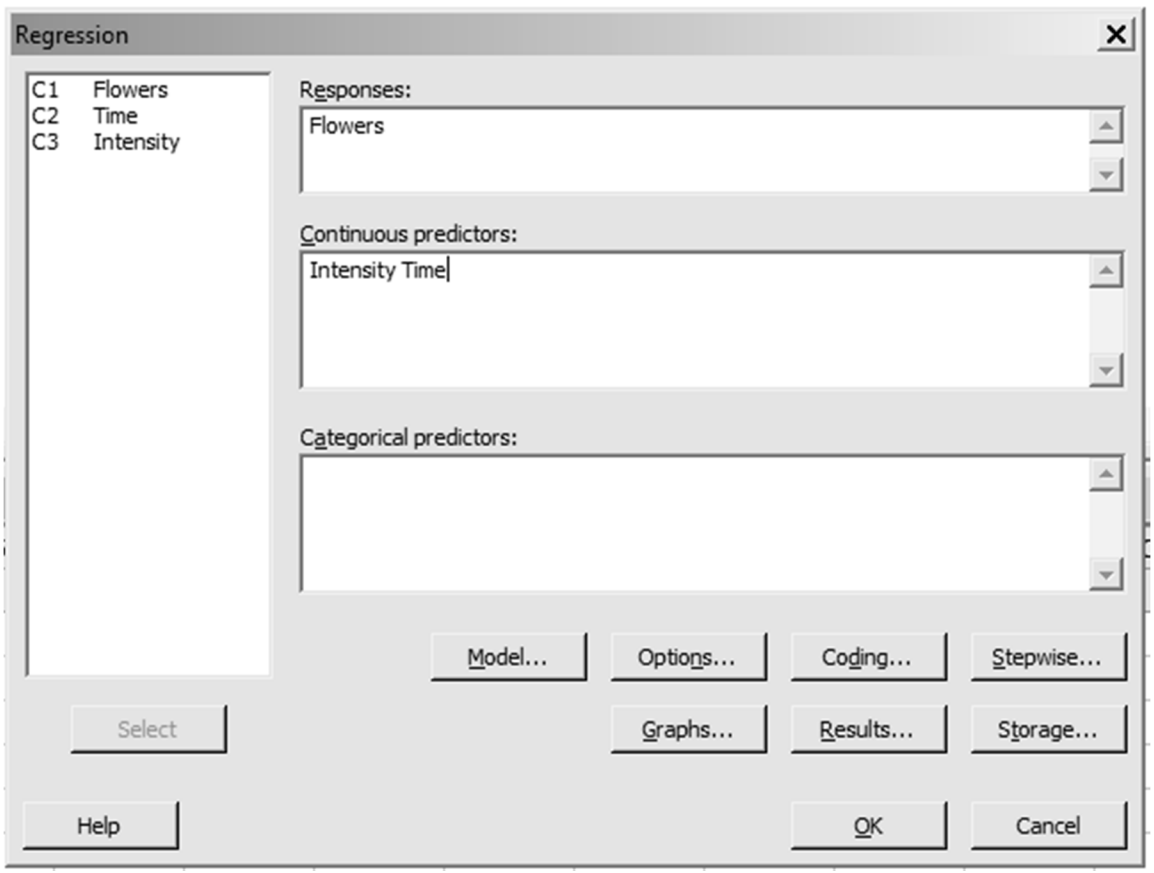
| Variable | INTENS | Maximum |
|---|---|---|
| FLOWERS | 150.00 | 77.40 |
| | 300.00 | 55.300 |
| | 450.00 | 61.90 |
| | 600.00 | 45.70 |
| | 750.00 | 44.90 |
| | 900.00 | 41.90 |

### Results for TIME = 2.00

| Variable | INTENS | N | N* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 |
|---|---|---|---|---|---|---|---|---|---|---|
| FLOWERS | 150.00 | 2 | 0 | 76.70 | 1.10 | 1.56 | 75.60 | * | 76.70 | * |
| | 300.00 | 2 | 0 | 73.55 | 4.45 | 6.29 | 69.10 | * | 73.55 | * |
| | 450.00 | 2 | 0 | 64.05 | 7.05 | 9.97 | 57.00 | * | 64.05 | * |
| | 600.00 | 2 | 0 | 57.55 | 5.35 | 7.57 | 52.20 | * | 57.55 | * |
| | 750.00 | 2 | 0 | 52.95 | 7.35 | 10.39 | 45.60 | * | 52.95 | * |
| | 900.00 | 2 | 0 | 48.50 | 4.10 | 5.80 | 44.40 | * | 48.50 | * |

```
Variable  INTENS  Maximum
FLOWERS   150.00    77.80
          300.00    78.00
          450.00    71.10
          600.00    62.90
          750.00    60.30
          900.00    52.60
```

**Step 3:** Fit a Multiple Linear Regression of FLOWER on two predictor variables, INTENS and TIME. To do this, Go to Stat → Regression; select FLOWERS into Response Variable and select INTENS and TIME into Predictors.

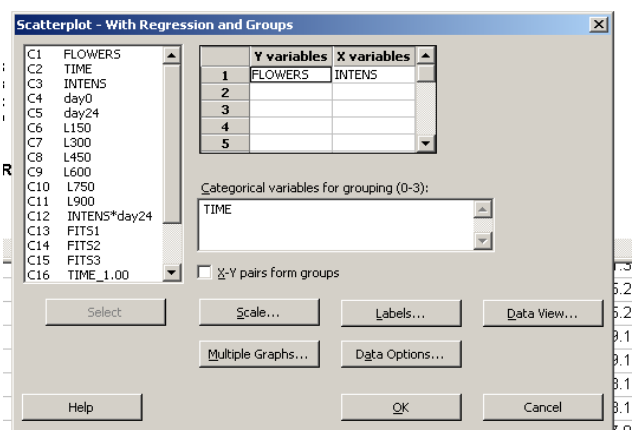


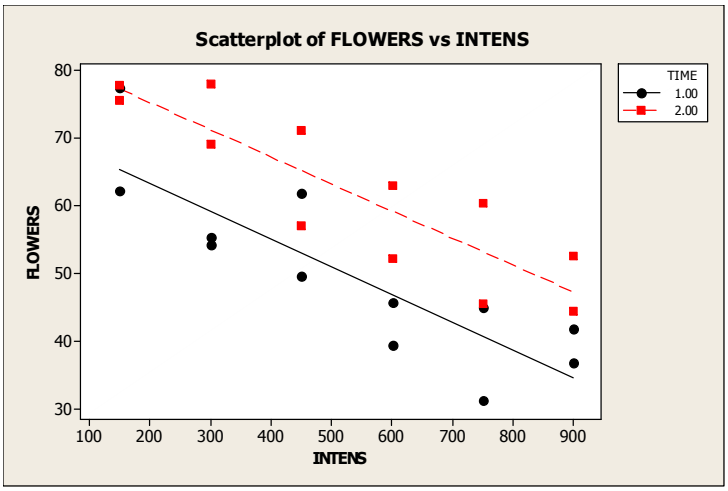

## Regression Analysis: Flowers versus Intensity, Time

```
Analysis of Variance

Source         DF  Adj SS   Adj MS  F-Value  P-Value
Regression      2  3466.7  1733.35    41.78    0.000
  Intensity     1  2579.8  2579.75    62.18    0.000
  Time          1   887.0   886.95    21.38    0.000
Error          21   871.2    41.49
  Lack-of-Fit   9   215.3    23.92     0.44    0.889
  Pure Error   12   655.9    54.66
Total          23  4337.9


Model Summary
```

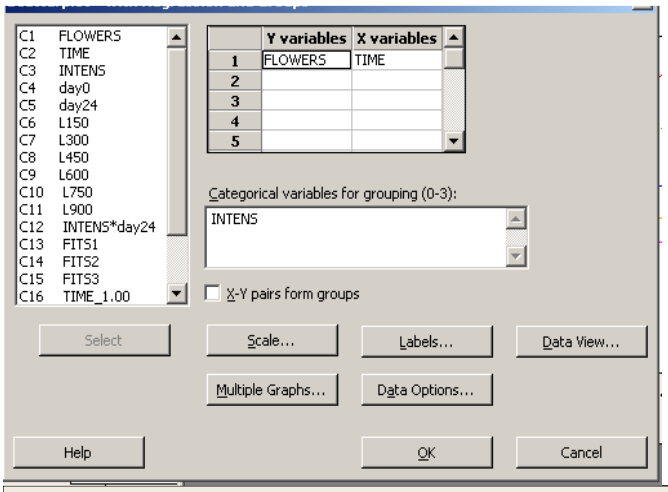
```
      S    R-sq  R-sq(adj)  R-sq(pred)
6.44107  79.92%     78.00%      73.84%


Coefficients

Term           Coef  SE Coef  T-Value  P-Value   VIF
Constant      59.15     4.95    11.94    0.000
Intensity  -0.04047  0.00513    -7.89    0.000  1.00
Time          12.16     2.63     4.62    0.000  1.00


Regression Equation

Flowers = 59.15 - 0.04047 Intensity + 12.16 Time


Fits and Diagnostics for Unusual Observations

                                Std
Obs  Flowers    Fit  Resid  Resid
  2    77.40  65.24  12.16   2.08  R

R  Large residual
```

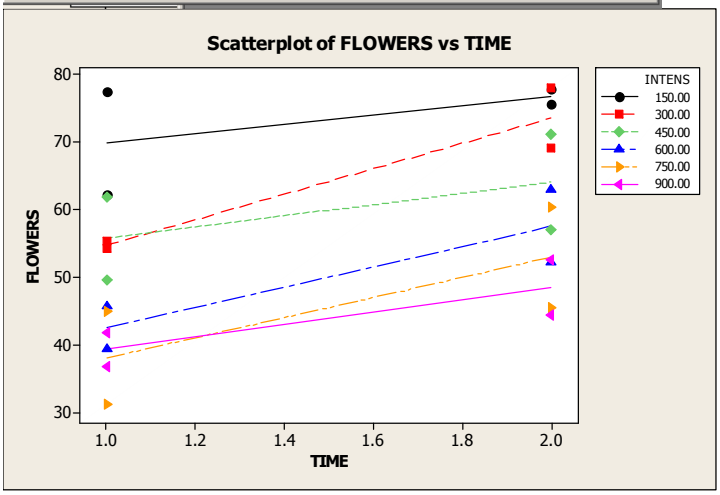
**Step 4: Scatterplot with Fitted Linear Regressions** for the two groups, viz., Early Timing and Late Timing. Go to Graphs → Scatterplot → With Regression and Groups; select FLOWERS into Y Variables and INTENS into X Variables; select TIME into Categorical Variables for Grouping, and click Ok to produce this graph. Also produce Residual plots.

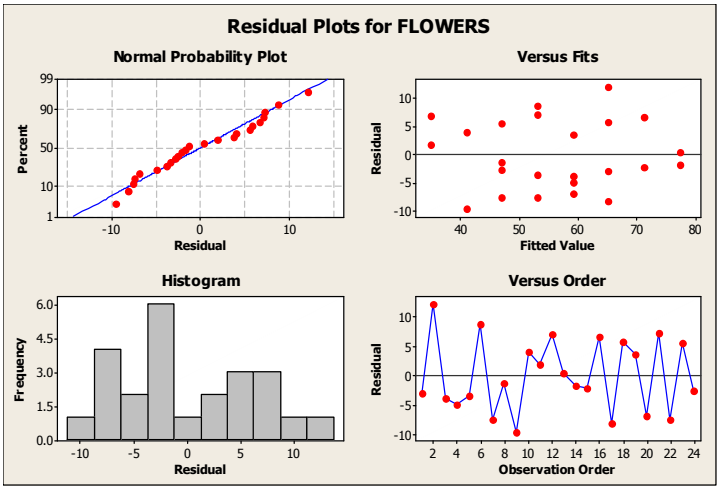




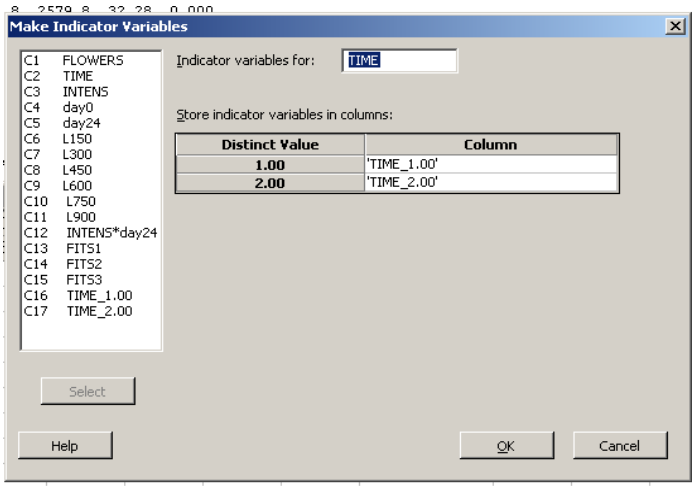

**Step 5: Scatterplot with Fitted Linear Regressions** for the two groups, viz., Early Timing and Late Timing. Go to Graphs → Scatterplot → With Regression and Groups; select FLOWERS into Y Variables and TIME into X Variables; select INTENS into Categorical Variables for Grouping, and click Ok to produce this graph.



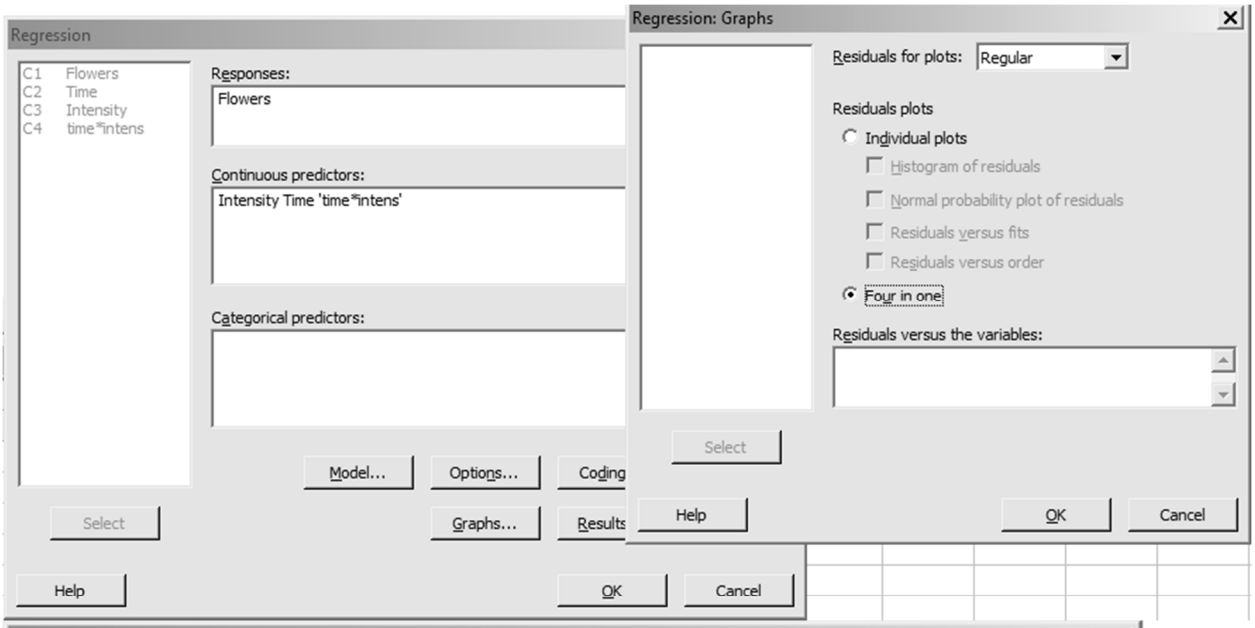



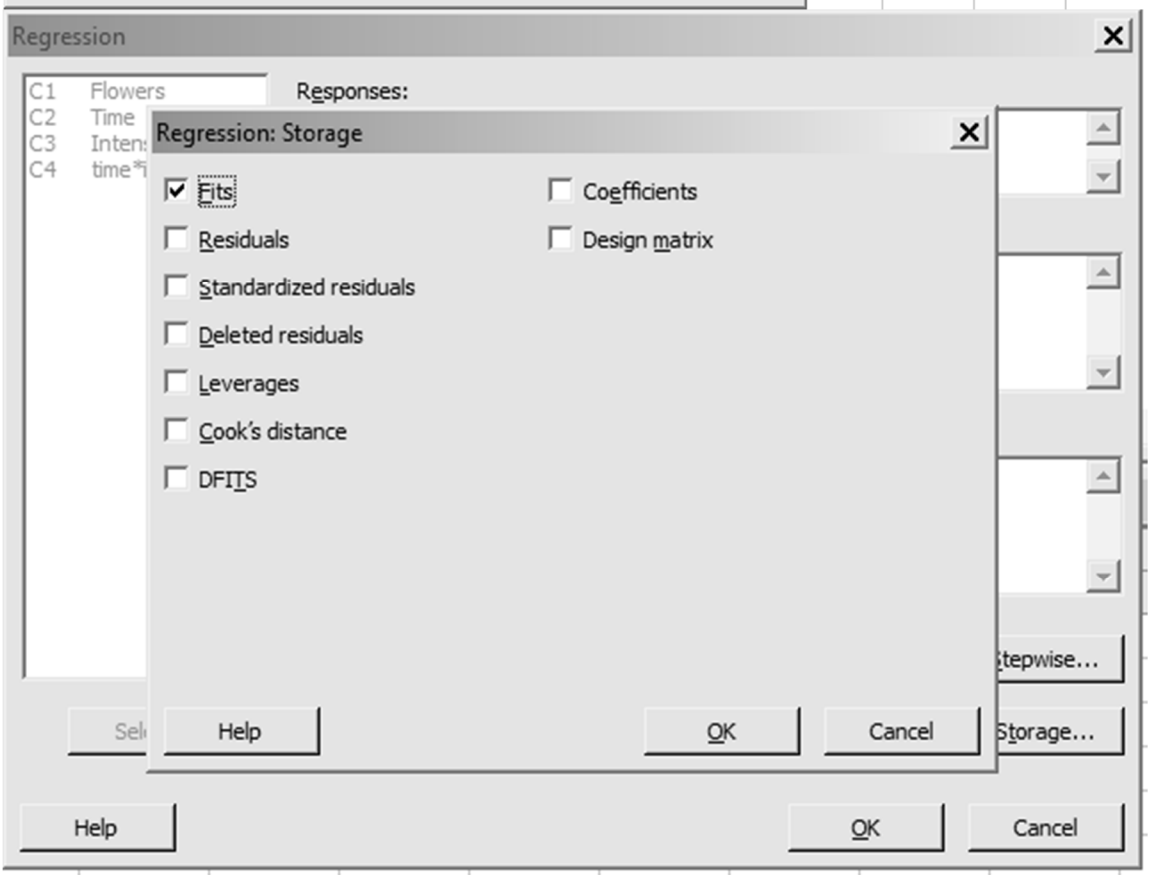



**Step 6**: **More Extensive way to carry out Multiple Regression,** with Categorical Predictors. Create Indicator Variables and Interaction Variables.
Create Indicator variables corresponding to the categorical variable TIME. Go to Calc → Make Indicator Variables, and select TIME. This creates two new columns C4 and C5, we rename them day0 and day24.
Similarly, we may also construct six indicator variables corresponding to INTENS, see ***R&S Display 9.7*** for the indicator variables L150 – L900.
Also, create an Interaction Variable between INTENS and day24 by going to Calc and calculating into Column C12 INTENS*day24.




**Step 7**: To run a regression of FLOWER on INTENS, day24 and INTENS*day24, go to STAT → Regression and select variables for analysis as shown below, and results follow. The fitted values are stored in FITS1 in column C13. To do this, click on Storage and click on Fits.




## Regression Analysis: Flowers versus Intensity, Time, time*intens

```
Analysis of Variance

Source          DF   Adj SS   Adj MS  F-Value  P-Value
Regression       3  3467.28  1155.76    26.55    0.000
```

| | | | | | |
|---|---|---|---|---|---|
| Intensity | 1 | 281.62 | 281.62 | 6.47 | 0.019 |
| Time | 1 | 153.22 | 153.22 | 3.52 | 0.075 |
| time*intens | 1 | 0.58 | 0.58 | 0.01 | 0.910 |
| Error | 20 | 870.66 | 43.53 | | |
| Lack-of-Fit | 8 | 214.73 | 26.84 | 0.49 | 0.841 |
| Pure Error | 12 | 655.92 | 54.66 | | |
| Total | 23 | 4337.94 | | | |

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 6.59795 | 79.93% | 76.92% | 70.95% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 60.10 | 9.71 | 6.19 | 0.000 | |
| Intensity | -0.0423 | 0.0166 | -2.54 | 0.019 | 10.00 |
| Time | 11.52 | 6.14 | 1.88 | 0.075 | 5.20 |
| time*intens | 0.0012 | 0.0105 | 0.12 | 0.910 | 14.20 |

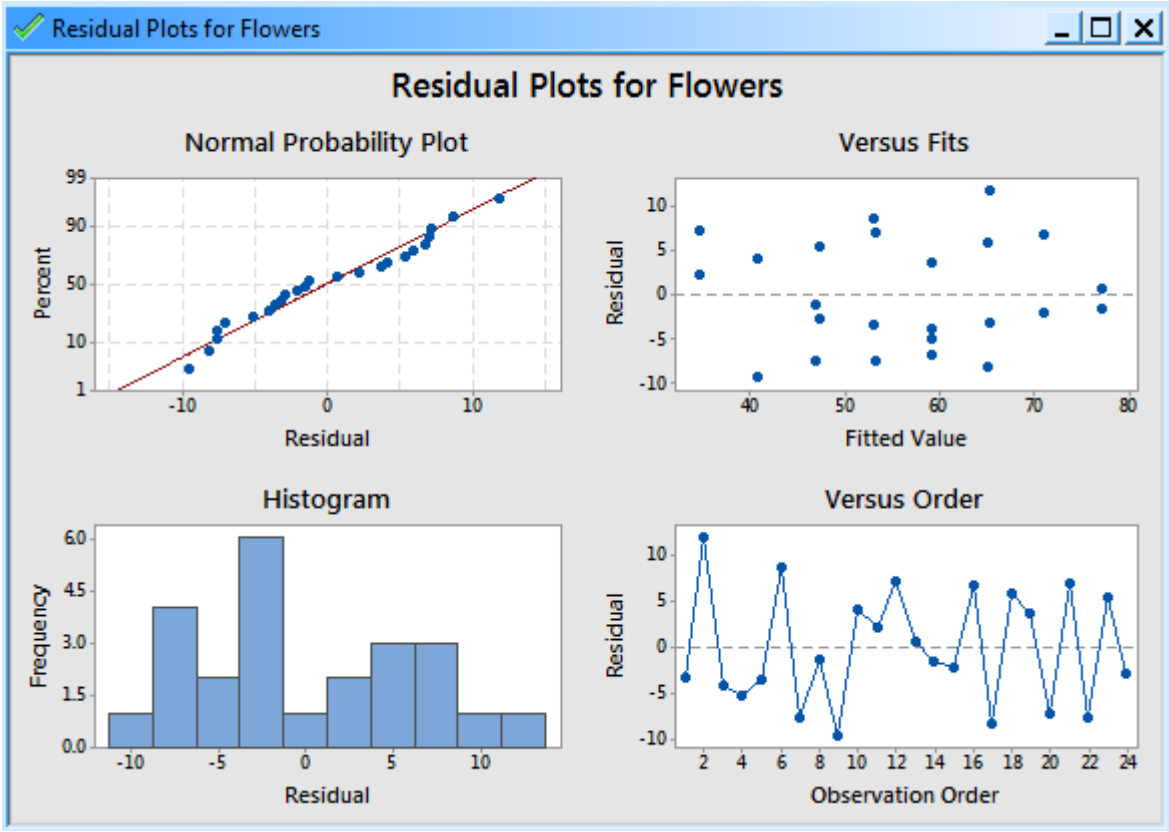
Regression Equation

Flowers = 60.10 - 0.0423 Intensity + 11.52 Time + 0.0012 time*intens

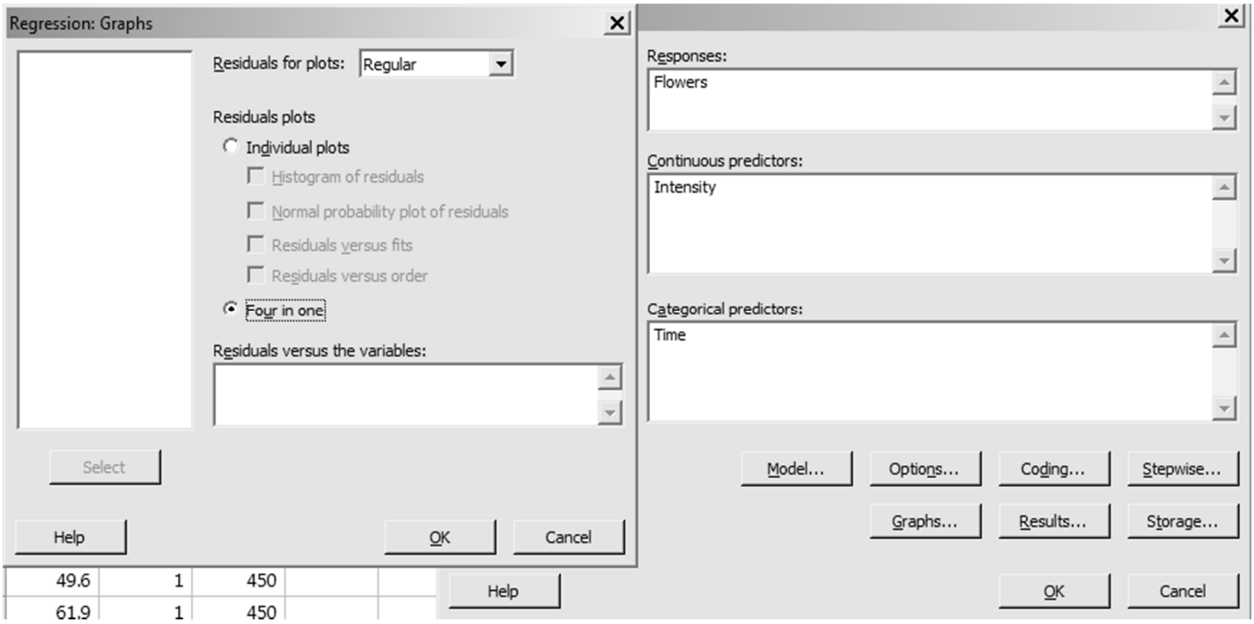
Fits and Diagnostics for Unusual Observations

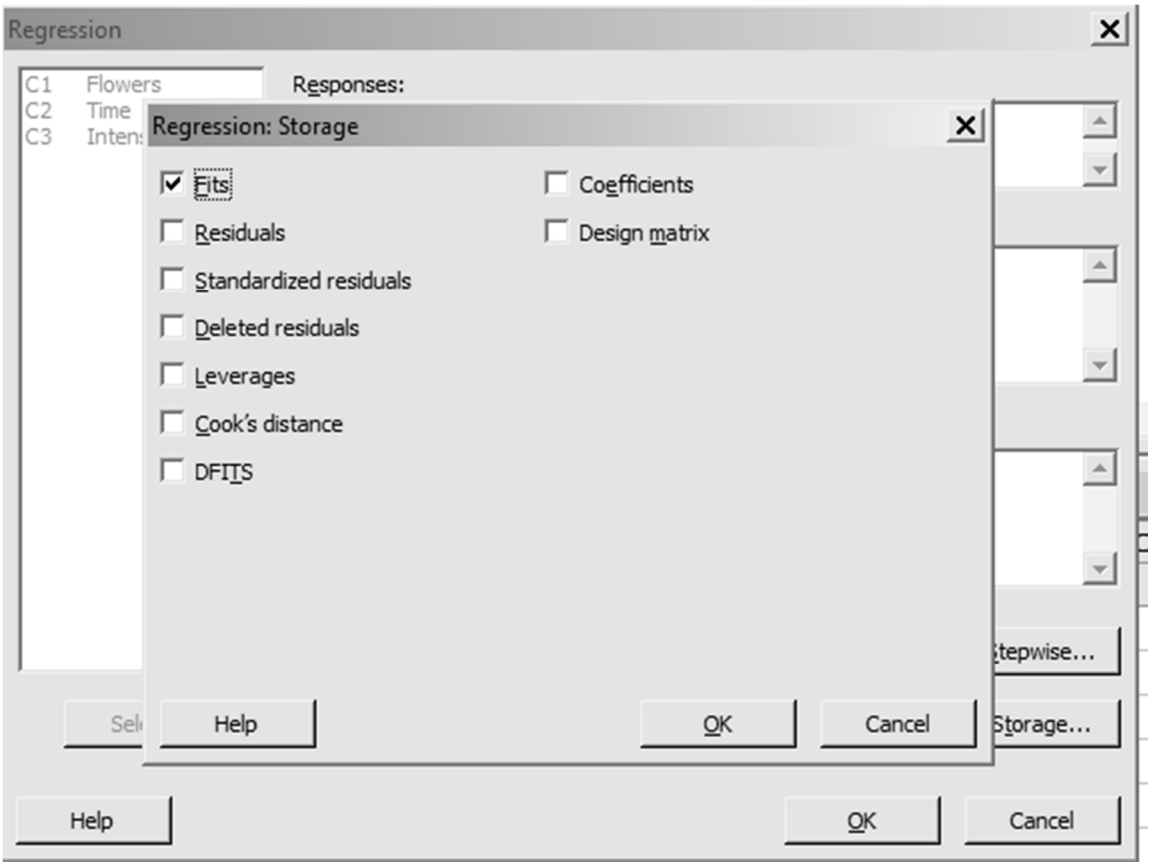
| Obs | Flowers | Fit | Resid | Std Resid | |
|---|---|---|---|---|---|
| 2 | 77.40 | 65.46 | 11.94 | 2.11 | R |

R  Large residual




Next, carry out a regression of FLOWER on INTENS and day24, but no interaction term. Output is shown below.

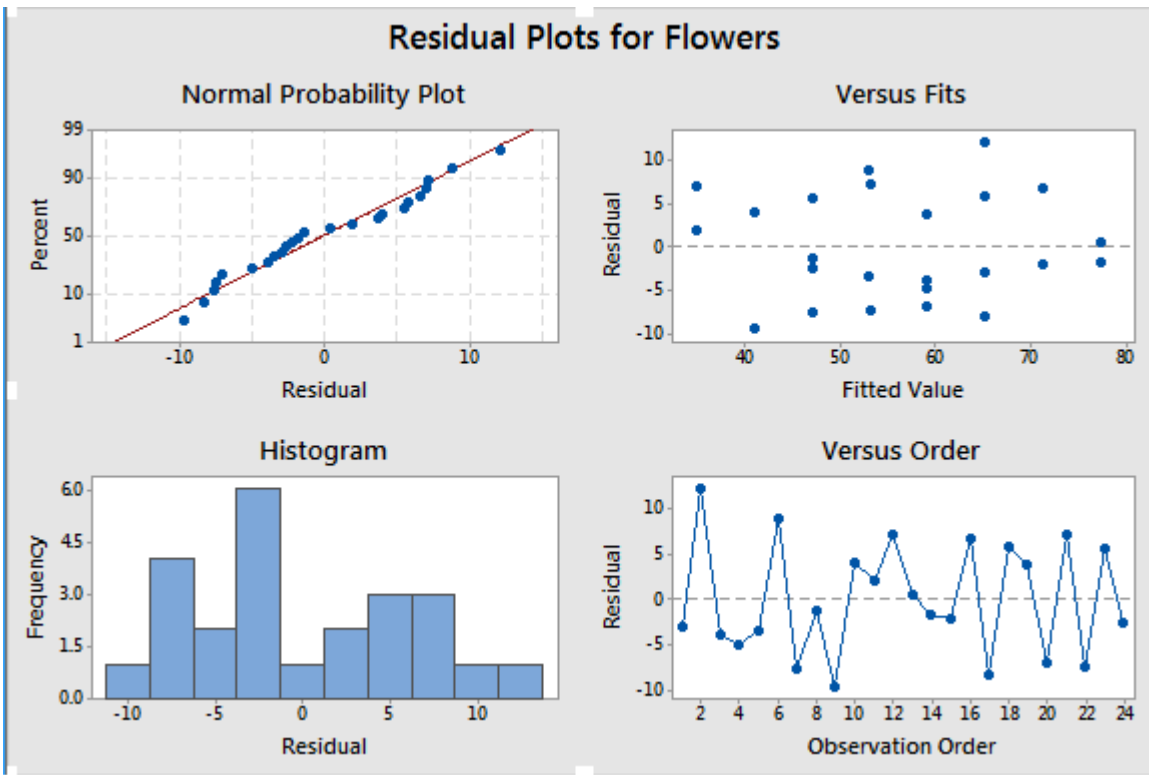




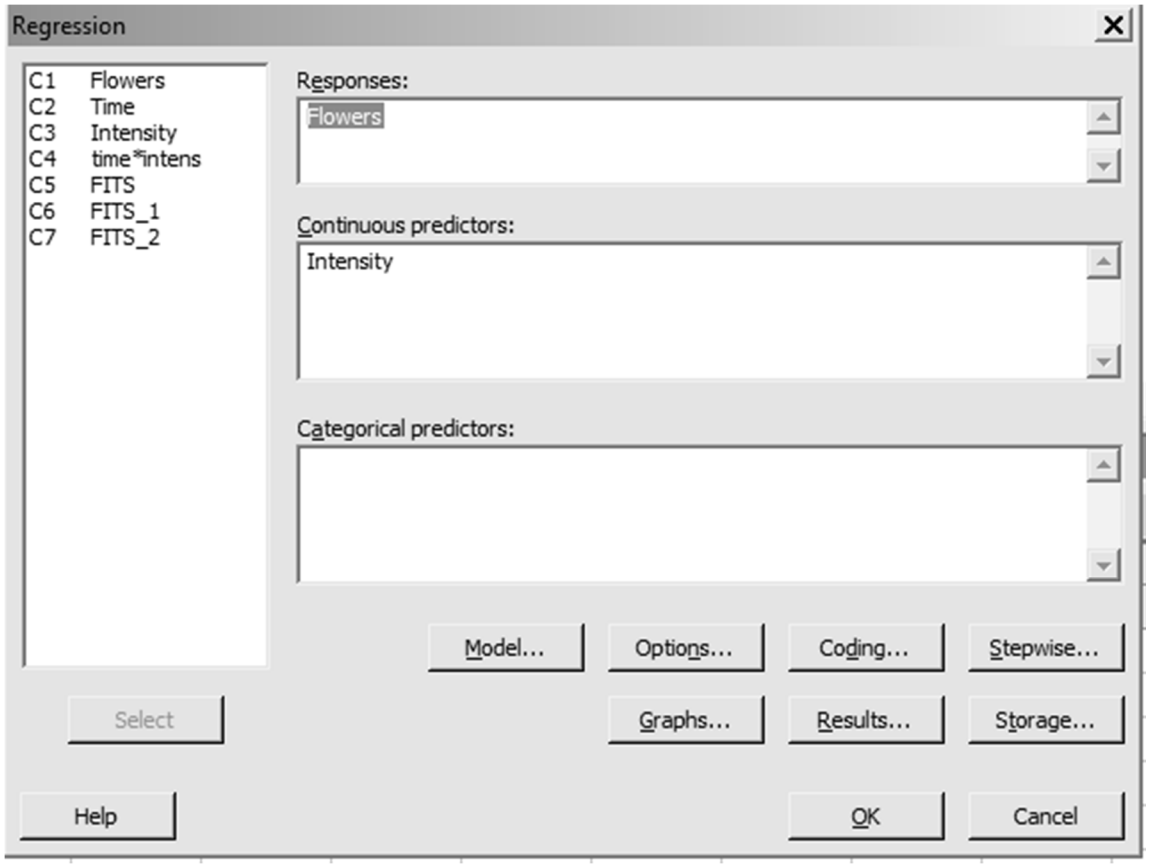

## Regression Analysis: Flowers versus Intensity, Time

```
Method

Categorical predictor coding  (1, 0)


Analysis of Variance

Source         DF  Adj SS   Adj MS  F-Value  P-Value
Regression      2  3466.7  1733.35    41.78    0.000
  Intensity     1  2579.8  2579.75    62.18    0.000
  Time          1   887.0   886.95    21.38    0.000
Error          21   871.2    41.49
  Lack-of-Fit   9   215.3    23.92     0.44    0.889
  Pure Error   12   655.9    54.66
Total          23  4337.9


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
6.44107  79.92%     78.00%      73.84%


Coefficients
```

```
Term           Coef  SE Coef  T-Value  P-Value   VIF
Constant      71.31     3.27    21.78    0.000
Intensity  -0.04047  0.00513    -7.89    0.000  1.00
Time
  2           12.16     2.63     4.62    0.000  1.00


Regression Equation

Time
1     Flowers = 71.31 - 0.04047 Intensity

2     Flowers = 83.46 - 0.04047 Intensity


Fits and Diagnostics for Unusual Observations

                                Std
Obs  Flowers    Fit  Resid  Resid
  2    77.40  65.24  12.16   2.08  R
```

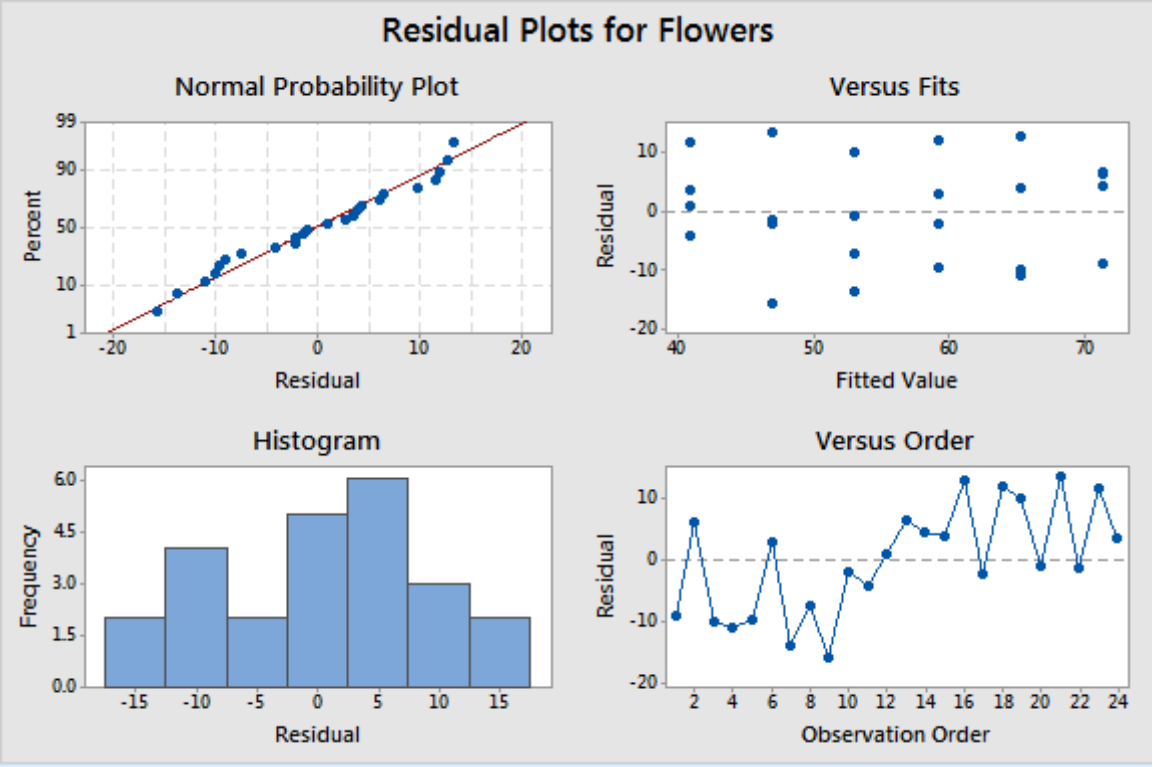



Finally, carry out a regression of FLOWER on INTENS only, output is shown below.




## Regression Analysis: Flowers versus Intensity

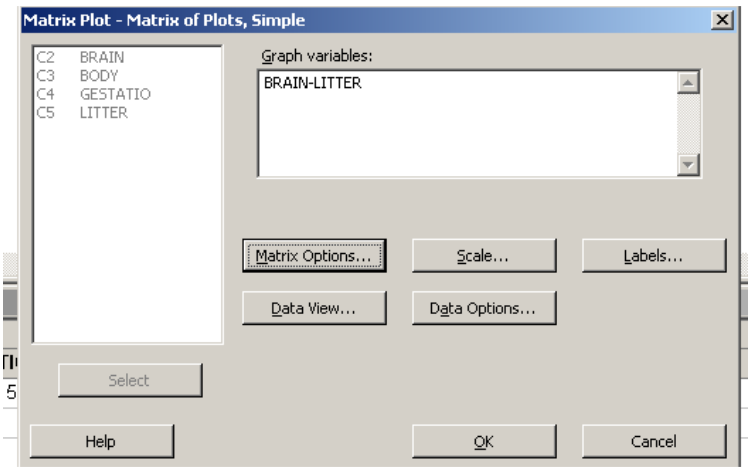
```
Analysis of Variance

Source         DF  Adj SS   Adj MS  F-Value  P-Value
Regression      1  2579.8  2579.75    32.28    0.000
  Intensity     1  2579.8  2579.75    32.28    0.000
Error          22  1758.2    79.92
  Lack-of-Fit   4   103.8    25.94     0.28    0.886
  Pure Error   18  1654.4    91.91
Total          23  4337.9


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
8.93966  59.47%     57.63%      52.42%


Coefficients

Term           Coef  SE Coef  T-Value  P-Value   VIF
Constant      77.39     4.16    18.60    0.000
Intensity  -0.04047  0.00712    -5.68    0.000  1.00


Regression Equation

Flowers = 77.39 - 0.04047 Intensity
```

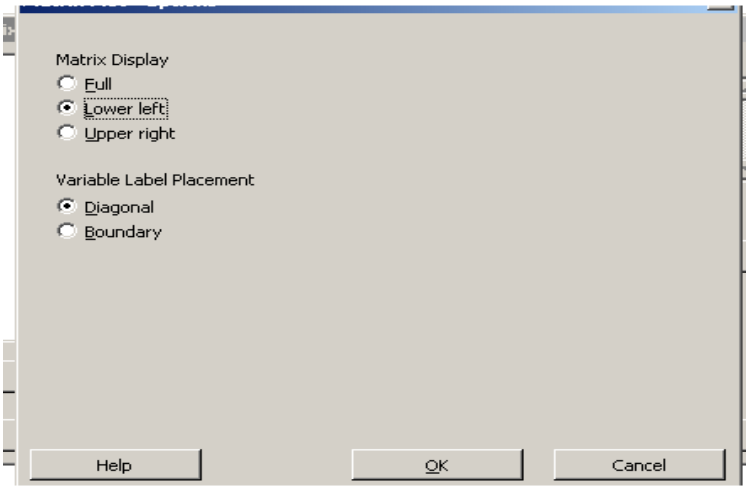

Residual Plots for Flowers
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Residual
Versus Order
Residual
Observation Order

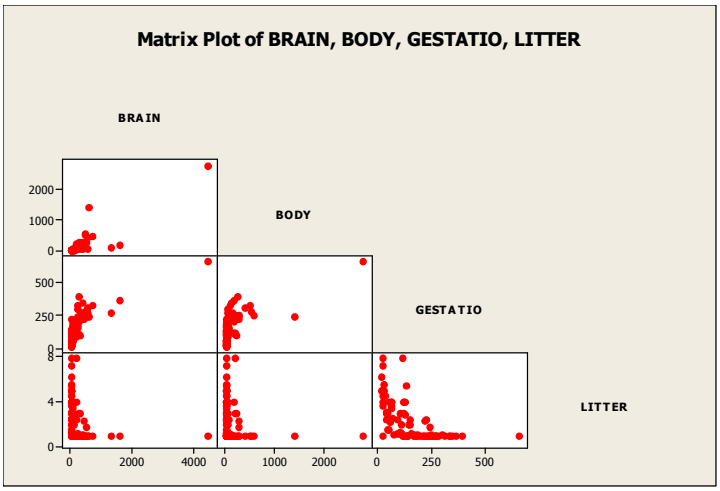
## Case 9.1.2. Why do some Mammals have Large Brains for their Size – An Observational Study. *R&S p.239-242.*

**Step 1: Copy the data into a Minitab Worksheet**: use these steps:
File → Open Worksheet →Browse your local directory and upload the csv file case0902.csv. To display the data in Minitab, go to Data →Display Data, and copy the columns C1-C5 in the window o n the right. The data will appear as five columns in Minitab. ***See R&S Display 9.4*** for data display on 96 different mammals.

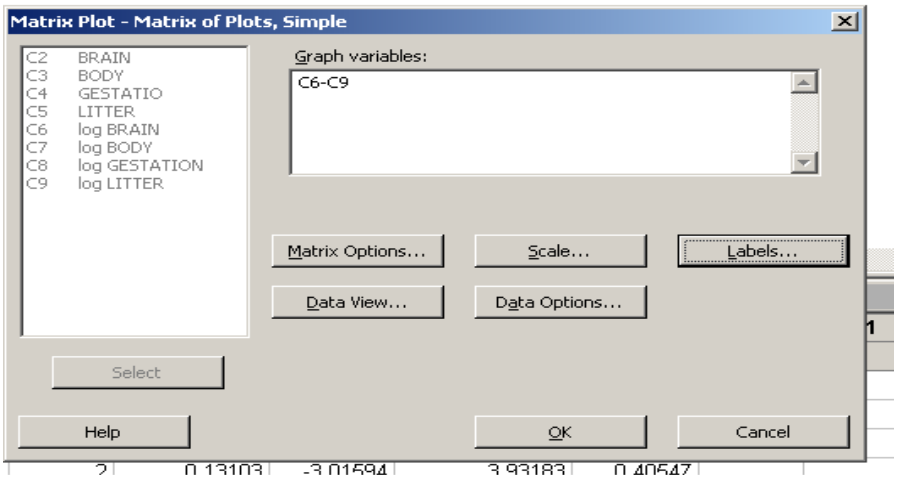
**Step 2: Go to Graph** → Matrix Plots → Simple; copy in Columns C2-C5 into the window on the right and click OK to get the plot below.

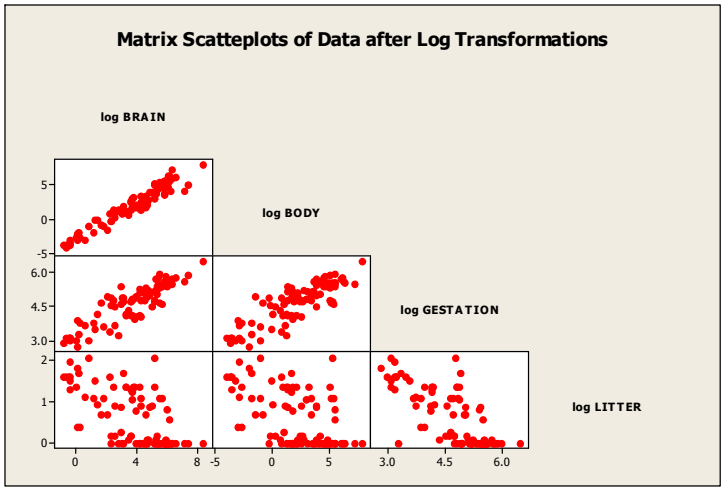


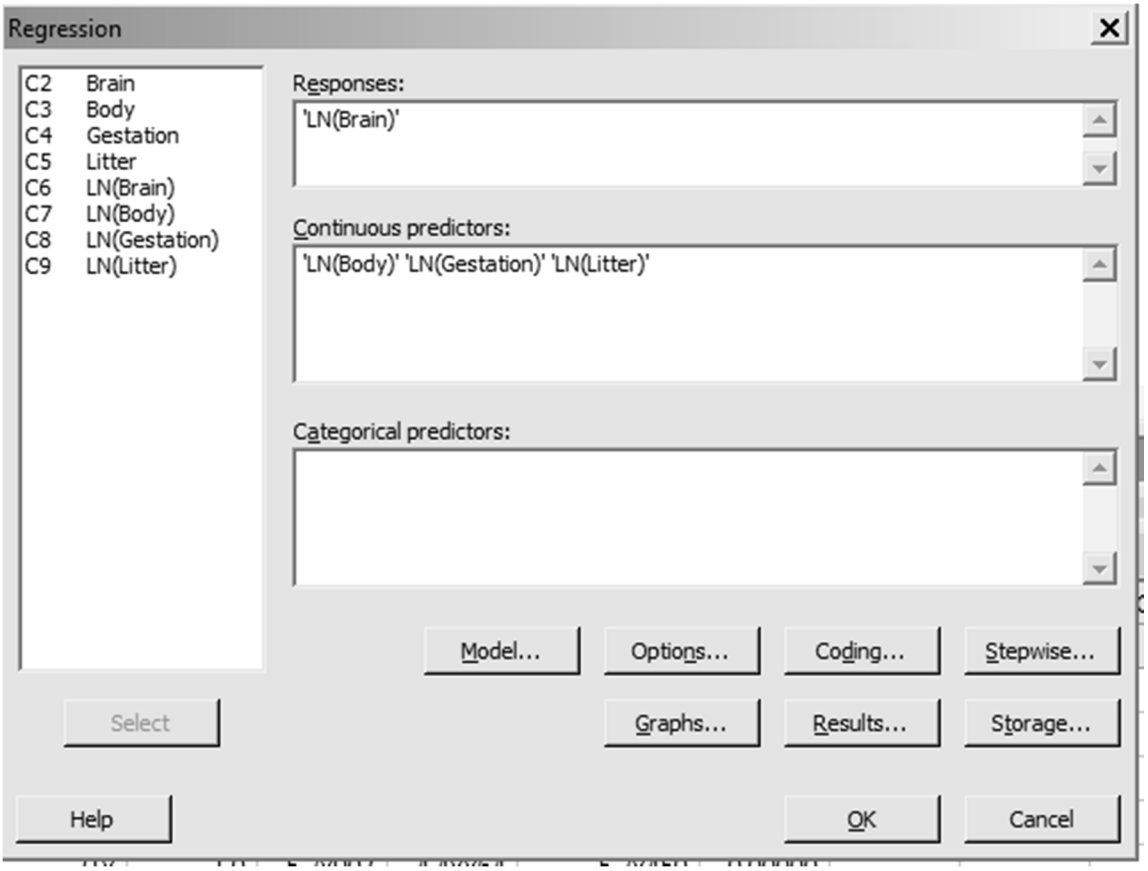



**Step 3:** Go to Calc → Calculator; and save into Columns C6-C9, the LN transformations of C2-C5. Then, go to Graph → Matrix Plots → Simple; copy in Columns C6-C9 into the window on the right and click OK to get the plot below.






**Step 4:** Go to Stat → Regression; select C6 log BRAIN into Response window and select C7- C9 into Predictors window; select Residuals Plots for graphs and click ok. See ***R&S Display 9.15***.



## Regression Analysis: LN(BRAIN) versus LN(Body), LN(Gestation), LN(Litter)

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 3 | 427.076 | 142.359 | 631.60 | 0.000 |
| LN(Body) | 1 | 70.189 | 70.189 | 311.41 | 0.000 |
| LN(Gestation) | 1 | 1.986 | 1.986 | 8.81 | 0.004 |
| LN(Litter) | 1 | 1.612 | 1.612 | 7.15 | 0.009 |
| Error | 92 | 20.736 | 0.225 | | |
| Total | 95 | 447.812 | | | |

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 0.474755 | 95.37% | 95.22% | 95.03% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 0.855 | 0.662 | 1.29 | 0.200 | |
| LN(Body) | 0.5751 | 0.0326 | 17.65 | 0.000 | 3.79 |
| LN(Gestation) | 0.418 | 0.141 | 2.97 | 0.004 | 6.27 |
| LN(Litter) | -0.310 | 0.116 | -2.67 | 0.009 | 2.54 |

Regression Equation

LN(BRAIN) = 0.855 + 0.5751 LN(Body) + 0.418 LN(Gestation) - 0.310 LN(Litter)

Fits and Diagnostics for Unusual Observations

| Obs | LN(BRAIN) | Fit | Resid | Std Resid | | |
|---|---|---|---|---|---|---|
| 25 | 7.378 | 6.233 | 1.144 | 2.45 | R | |
| 28 | 5.193 | 5.211 | -0.018 | -0.04 | | X |
| 48 | 6.380 | 7.311 | -0.931 | -2.03 | R | |
| 53 | 7.170 | 5.595 | 1.575 | 3.36 | R | |
| 72 | 2.862 | 2.937 | -0.075 | -0.18 | | X |
| 86 | 5.521 | 6.476 | -0.954 | -2.05 | R | |

R  Large residual
X  Unusual X