Robert Brandl - I pledge my honor that I have abided by the Stevens Honor System.

Final Project

1 Report

1.1 Executive Summary

As part of this project, I have investigated the relationship between different variables that affect bones. This analysis focuses on the processes of bone formation, measured through the variable VO+, and bone resorption, measured through the variable VO-. Additionally, the variables Osteocalcin (OC), a biochemical marker for bone formation, and tartrate-resistant acid phosphatase (TRAP), a biochemical marker for bone resorption, were utilized. The goal of the project was to determine the best model using regression that can predict VO+ and VO-, using the other variables to predict these variables and predict the amount of bone formation and bone resorption.

1.2 Data Set

For this analysis, the biomark.xls data set was used, which contains the values for the VO+, VO-, OC, and TRAP variables, with the following respective headings, “voplus,” “vominus,” “voc,” and “vtrap”. For the OC variable, measurements were taken in milligrams of OC per milliliter of blood (mg/ml), and for TRAP, measurements were taken in units per liter (U/l). This information was generated from a study consisting of 31 healthy women aged 11 to 32 years old. In addition, the logarithmic versions of each variable are also contained within the data set to use for further analysis, where these variables were titled “lvoplus,” “lvominus,” “loc,” and “ltrap”.

1.3 Methods of Statistical Analysis

To perform an analysis of the data set, different simple and multiple linear regression models were generated. The simple linear regression models utilize only one explanatory variable to predict another variable, while the multiple linear regression models use more than one variable to predict one variable. These models were generated using a varying combination of VO+, VO-, OC, and TRAP variables to predict either VO+ or VO- accordingly. In addition, these models were also generated using the logarithmic versions of these variables.

1.4 Software

To generate numerical and graphical data and statistical models for the dataset, the R programming language, with its builtin packages, was used through the development environment of RStudio Version 1.4.1717. The report and compilation of data were generated using Google Docs to allow for text editing and report formatting.

1.5 Model Analysis and Conclusions

To better understand the data, numerical and graphical summaries were generated for each variable, as well as the relationships between the different variables. This included producing the mean, median, standard deviation and scatter plots for each variable, which indicated that the data was normally distributed with minimal outliers or unusual patterns. The correlations between the variables, both non-logarithmic and logarithmic, were as follows:

```
           voplus   vominus        oc      trap
voplus  1.0000000 0.8957707 0.6596140 0.7648649
vominus 0.8957707 1.0000000 0.4547603 0.6779267
oc      0.6596140 0.4547603 1.0000000 0.7298519
trap    0.7648649 0.6779267 0.7298519 1.0000000

          lvoplus  lvominus       loc     ltrap
lvoplus  1.0000000 0.8396741 0.7735853 0.7549684
lvominus 0.8396741 1.0000000 0.5546070 0.6643005
loc      0.7735853 0.5546070 1.0000000 0.7953528
ltrap    0.7549684 0.6643005 0.7953528 1.0000000
```

Using these correlations, multiple regression models were created to predict VO+, VO-, LVO+, and LVO-. For each response variable, one simple regression model and three multiple linear regression models were created. Overall, for each response variable, the simple linear regression model failed to produce the most accurate predictions, with residuals that often appeared slightly curved with some outliers. However, the final two multiple linear regression models produced the most accurate results for the predictions, measured through the significance of the coefficients of the model, the residual standard error or sigma value, and the percent of variation or R-Squared value. These values can be seen with more detail and with simple descriptions in the second section, Data, below.

After creating multiple models, the best model for predicting VO+ was the multiple linear regression model, part of exercise 11.38 (e) in the Data section, using OC and VO- as explanatory variables. This model produced the greatest percent of variation of R-Squared value

of 0.8742 or 87.42%, higher than the other three models, as well as the lowest sigma or residual standard error value of 205.6. The best model for predicting VO- was the multiple linear regression model using OC, TRAP, and VO+ as explanatory variables, displayed in exercise 11.40 in the Data section. This model produced the greatest R-Squared value of 0.8245 or 82.45% and had the lowest residual standard error value of 179.2.

For predicting LVO+, the best model was the multiple linear regression model using LOC and LVO- as explanatory variables, seen in exercise 11.39 in the Data section. This model produced the greatest R-Squared value of 0.8307 or 83.07% and had the lowest residual standard error value of 0.2286. For predicting LVO-, the best model was the multiple linear regression model using LVO+, LOC, and LTRAP as explanatory variables, shown in exercise 11.41. This model produced the greatest R-Squared value of 0.7197 or 71.97% and had the lowest residual standard error value of 0.2558.

All in all, the multiple linear regression models produced more accurate and significant results than the simple linear regression model for generating prediction. The logarithmic variables also demonstrated certain advantages in drawing conclusions, specifically the more significant p-values which resulted from the models and the higher accuracy of predictions. For further understanding of the data, see the extended Data section below.

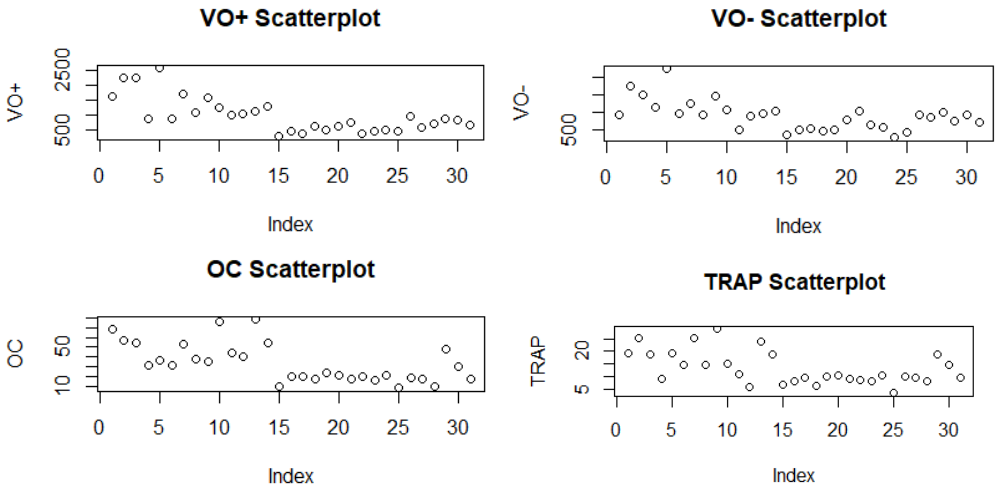
2 Data

11.36

a) Numerical Summary for VO+, VO-, OC, and TRAP

	VO+	VO-	OC	TRAP
Mean	985.80	889.20	33.42	13.25
Standard Deviation	579.86	427.62	19.61	6.53
Minimum	285.00	254.00	8.10	3.30
Maximum	2545.00	2236.00	77.90	28.80
First Quartile	542.50	554.00	18.60	8.90
Median	870.00	903.00	30.20	10.30
Third Quartile	1188.50	1023.00	46.05	18.80

Graphical Summaries:



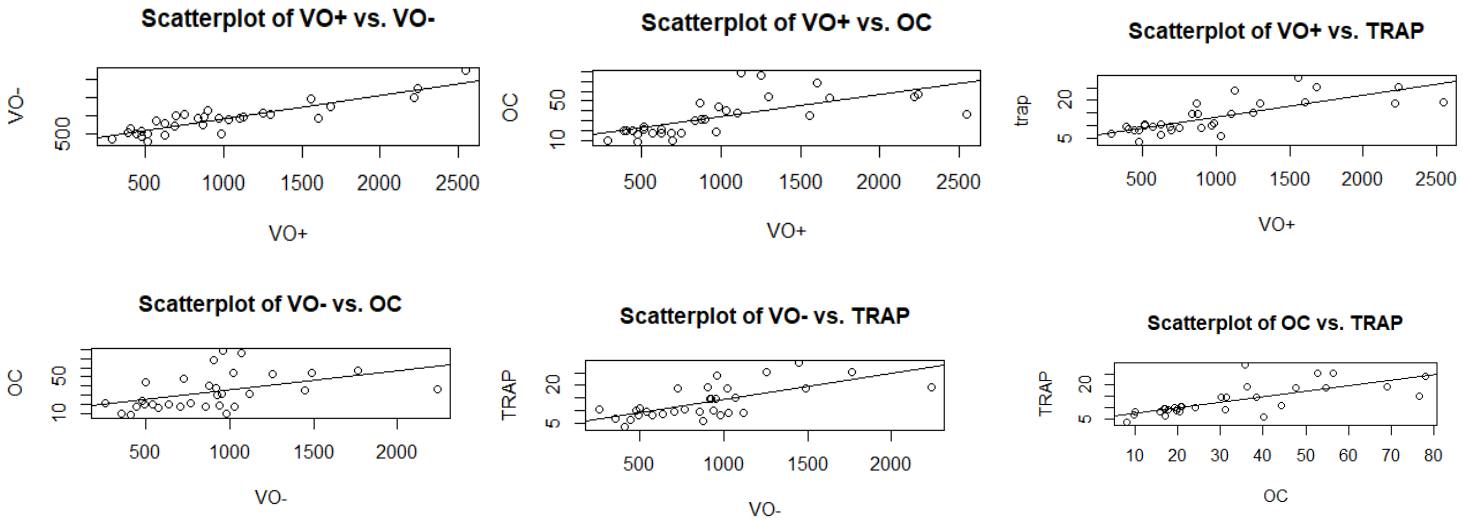
The distributions for all of the variables appear scattered, with notably high maximum values for the VO+ and VO- variables.

b)

Numerical Summary of Relationship Between Variables Using Correlation (calculated using Pearson's correlation coefficient):

	voplus	vominus	oc	trap
voplus	1.0000000	0.8957707	0.6596140	0.7648649
vominus	0.8957707	1.0000000	0.4547603	0.6779267
oc	0.6596140	0.4547603	1.0000000	0.7298519
trap	0.7648649	0.6779267	0.7298519	1.0000000

Graphical Summary (with best fit lines):



From the results of the numerical and graphical summaries, it is clear that the relationships between each pair of variables are linear and positively correlated. However, each pair of variables has a different strength of correlation. The relationships between VO+ & VO-, VO+ & TRAP, and OC & TRAP are all strong positive linear associations. The relationships between VO+ & OC and VO- & TRAP are also positive associations but weaker than the previous relationships. The final pair of variables, VO- & OC, shows the weakest positive association, with a correlation coefficient of 0.45 and a flatter sloped line in the scatter plot.

11.37

a) Using simple linear regression model to predict VO+ using OC:

Residual Values Summary:

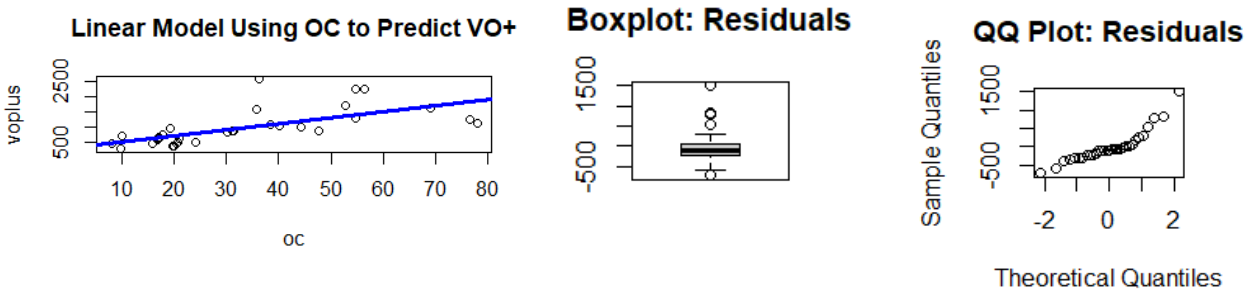
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-727.45	-234.43	-85.08	43.66	1500.99

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  334.034    159.241   2.098   0.0448 *
oc            19.505      4.127   4.726 5.43e-05 ***
```

Residual Standard Error: 443.3 with degrees of freedom = 29

Multiple R-squared: 0.4351, Adjusted R-squared: 0.4156

F-statistic: 22.34 with degrees of freedom = 1, 29 → P-value: 5.429e-05



The scatter plot generated above displays greater variation in VO+ for large values of OC. Model generated: VO+ = 334 + 19.51(OC), which displays a t-statistic of 4.73, where the resulting p-value < 0.0005, indicating significant results. Therefore, there is enough evidence that OC can be used as an explanatory variable to predict VO+. The generated plot of residuals and

QQ plot using OC indicates the residuals are slightly curved and somewhat deviate from the normal distribution.

b) Using multiple linear regression model to predict VO+ using OC and TRAP:

Residual Values Summary:

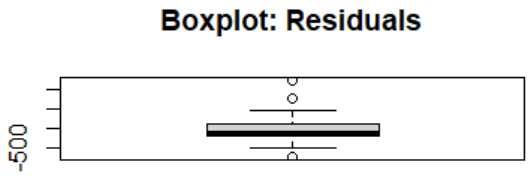
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-708.2	-198.6	-100.2	125.8	1224.8

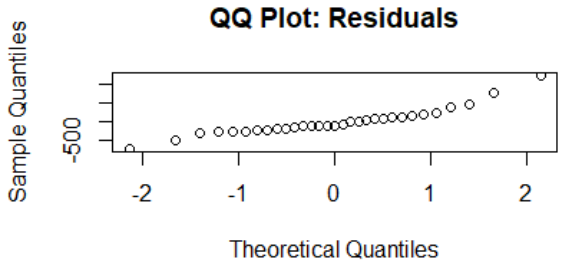
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   57.704    156.539   0.369  0.71518
oc             6.415      5.125   1.252  0.22102
trap          53.874     15.393   3.500  0.00158 **
```

Residual Standard Error: 376.3 with degrees of freedom = 28

Multiple R-squared: 0.607, Adjusted R-squared: 0.5789

F-statistic: 21.62 with degrees of freedom = 2, 28 → P-value: 2.096e-06





Model: VO+ = 57.7 + 6.42(OC) + 53.87(TRAP).

The coefficient of OC is not statistically significant, since the t-value = 1.25, resulting in a p-value of 0.221, which is much greater than the significance level of 0.05. However, the coefficient of TRAP is statistically significant, with a t-value of 3.50 and p-value of 0.002, indicating that TRAP is a better predictor of bone formation, VO+. This is consistent with the correlations found in Exercise 11.36.

11.38

a) Model: VO+ = $\beta_0 + \beta_1$ (OC) + β_2 (TRAP) + β_3 (VO-) + ε_i, where ε_i are independent N(0,σ) variables

b) Using multiple linear regression model to predict VO+ using OC, TRAP, and VO-:

Residual Values Summary:

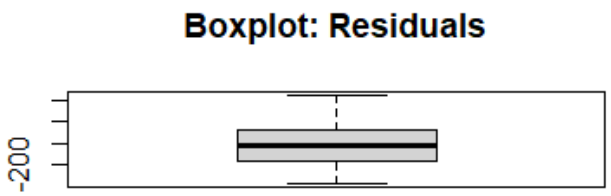
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-364.19	-158.57	-15.13	120.08	441.11

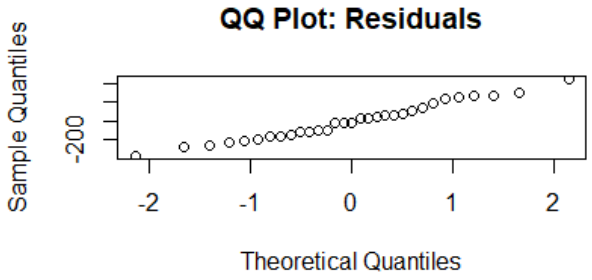
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -243.4877    94.2183  -2.584  0.01549 *
oc             8.2349     2.8397   2.900  0.00733 **
trap           6.6071    10.3340   0.639  0.52797
vominus        0.9746     0.1211   8.048  1.2e-08 ***
```

Residual Standard Error: 207.8 with degrees of freedom = 27

Multiple R-squared: 0.884, Adjusted R-squared: 0.8715

F-statistic: 68.84 with degrees of freedom = 3, 27 → P-value: 9.031e-13





Since the results demonstrate p-values for the coefficients of OC and VO- below the significance level of 0.05, these variables represent strong explanatory variables in predicting VO+. However, the p-value for TRAP is 0.528, greater than the significance level, and does not produce significant results. Clearly, the best explanatory variable in predicting VO+ is VO-.

c)

Model			Linear Regression (11.37 (a))	Multiple Linear Regression (11.37 (b))	Multiple Linear Regression (11.38 (b))
	OC	Estimated Coefficient	19.51	6.42	8.23
		Standard Error	4.13	5.13	2.84
		t statistic	4.73	1.25	2.90
		p-value	5.43e-05	0.22	0.0073

Explanatory Variable	TRAP	Estimated Coefficient	N/A	53.87	6.61
		Standard Error		15.39	10.33
		t statistic		3.50	0.64
		p-value		0.0016	0.53
	VO-	Estimated Coefficient	N/A	N/A	0.97
		Standard Error			0.12
		t statistic			8.05
		p-value			1.2e-08

The coefficient for the OC variable, which was used in all three regression models, changed for each model. However, the p-value appears significant for the first and third models, where the values are less than the significance level of 0.05, indicating that the variable acts as a strong explanatory variable for predicting VO+ in a model with just OC or with OC & TRAP & VO-. The coefficient for the TRAP variable, used in the second and third models, is significantly different between the two models, with a significant p-value in the second model but not the third. This indicates that the TRAP variable acts as a strong explanatory variable in a model with OC & TRAP, but not with OC & TRAP & VO-. Finally, the coefficient for the VO- variable was only used in the third model, with a highly significant p-value, demonstrating that VO- can be used to predict VO+ when used in conjunction with OC & TRAP.

d)

Model	Linear Regression (11.37 (a))	Multiple Linear Regression (11.37 (b))	Multiple Linear Regression (11.38 (b))
Percent of Variation (adjusted R-Squared)	0.4156 → 41.56%	0.5789 → 57.89%	0.8715 → 87.15%
Estimate of σ (residual standard error)	443.3	376.3	207.8

The coefficient of determination, or the percent of variation, indicates the performance of the regression model. Therefore, the final multiple linear regression model exhibits the greatest

R-Squared value, demonstrating that that model performs the best. Furthermore, the estimate of σ is the lowest for the final model, indicating the lowest residual standard error of the 3 models.

e) Using multiple linear regression model to predict VO+ using OC and VO-:

Residual Values Summary:

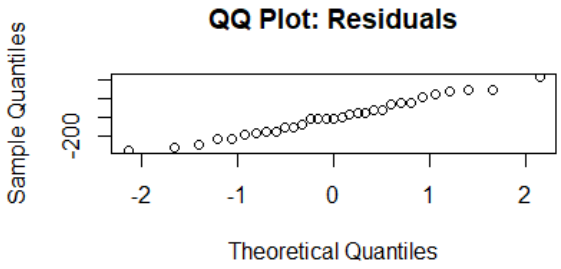
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-350.25	-153.94	-13.22	148.19	428.09

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -234.14400   92.09009  -2.543 0.016818 *
oc             9.40388    2.14964   4.375 0.000153 ***
vominus        1.01857    0.09858  10.333 4.65e-11 ***
```

Residual Standard Error: 205.6 with degrees of freedom = 28

Multiple R-squared: 0.8826, Adjusted R-squared: 0.8742

F-statistic: 105.30 with degrees of freedom = 2, 28 → P-value: 9.418e-14



Since the results demonstrate p-values for the coefficients of OC and VO- below the significance level of 0.05, these variables represent strong explanatory variables in predicting VO+. These results show even more significant p-values for the OC and VO- variables with the exclusion of the TRAP variable compared to part (b).

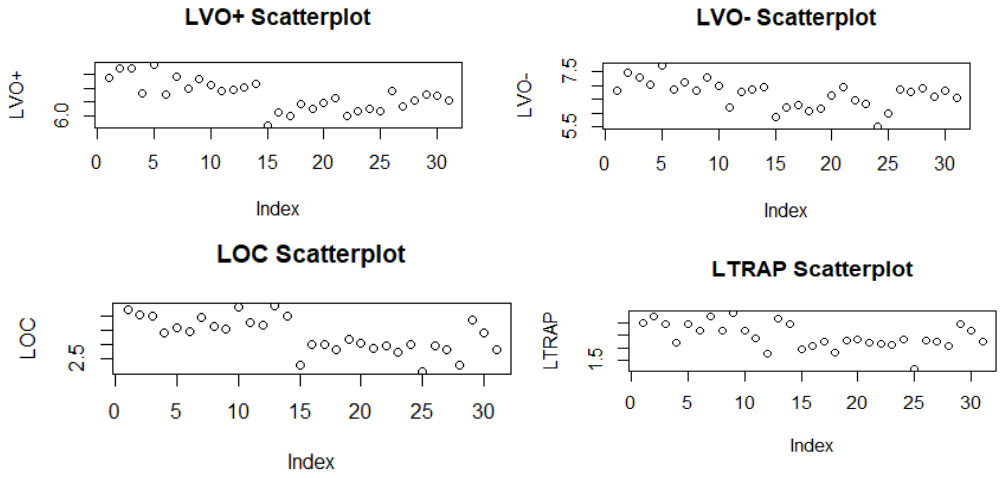
11.39

Numerical Summary for LVO+, LVO-, LOC, and LTRAP

	LVO+	LVO-	LOC	LTRAP
Mean	6.74	6.68	3.34	2.47
Standard Deviation	0.56	0.48	0.61	0.50

Minimum	5.65	5.54	2.09	1.19
Maximum	7.84	7.71	4.36	3.36
First Quartile	6.30	6.32	2.92	2.19
Median	6.77	6.81	3.41	2.33
Third Quartile	7.08	6.93	3.83	2.93

Graphical Summaries:

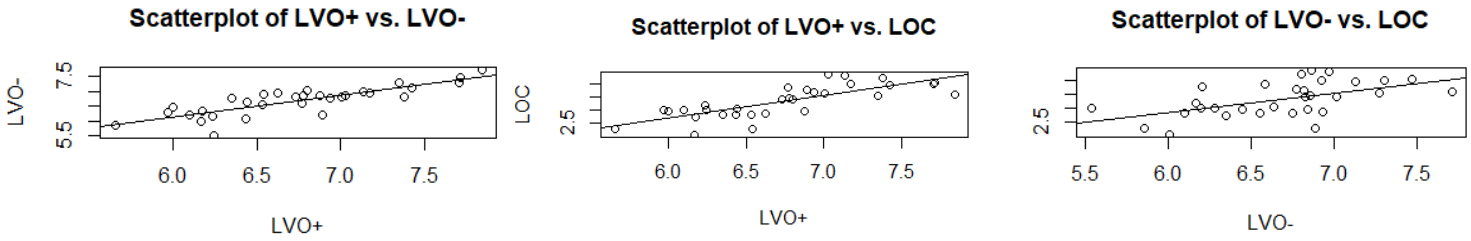


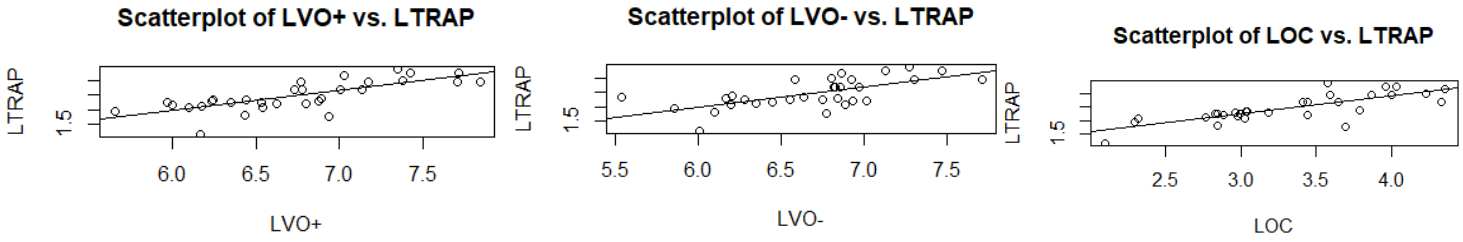
The distributions for all of the variables appear scattered, with less extreme outliers or variation compared to the non-logarithmic variables.

Numerical Summary of Relationship Between Variables Using Correlation (calculated using Pearson's correlation coefficient):

```
          lvoplus  lvominus       loc     ltrap
lvoplus  1.0000000 0.8396741 0.7735853 0.7549684
lvominus 0.8396741 1.0000000 0.5546070 0.6643005
loc      0.7735853 0.5546070 1.0000000 0.7953528
ltrap    0.7549684 0.6643005 0.7953528 1.0000000
```

Graphical Summary (with best fit lines):





From the results of the numerical and graphical summaries, it is clear that the relationships between each pair of variables are linear and positively correlated. However, each pair of variables has a different strength of correlation. The relationships between LVO+ & LVO-, LVO+ & LTRAP, and LOC & LTRAP are all strong positive linear associations. The relationships between LVO+ & LOC and LVO- & LTRAP are also positive associations but weaker than the previous relationships. The final pair of variables, LVO- & LOC, shows the weakest positive association, with a correlation coefficient of 0.55 and a flatter sloped line in the scatter plot, but this correlation is stronger than the non-logarithmic versions of the variables.

Using simple linear regression model to predict LVO+ using LOC:

Residual Values Summary:

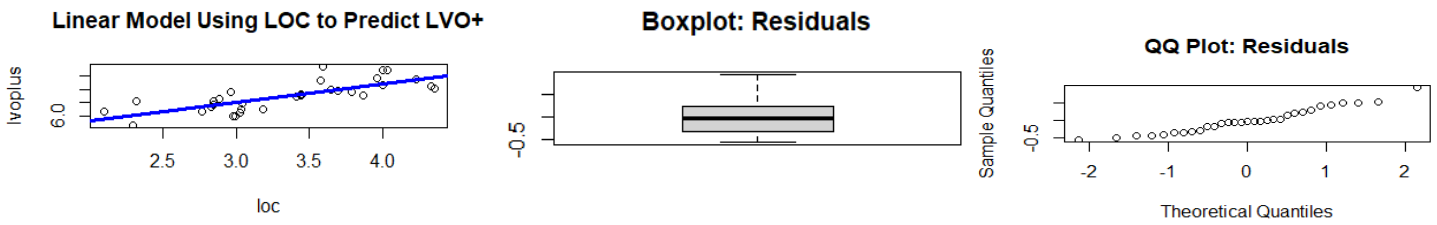
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.53119	-0.30174	-0.03742	0.22220	0.91892

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.3846     0.3643  12.036 8.42e-13 ***
loc           0.7062     0.1074   6.574 3.34e-07 ***
```

Residual Standard Error: 0.358 with degrees of freedom = 29

Multiple R-squared: 0.5984, Adjusted R-squared: 0.5846

F-statistic: 43.22 with degrees of freedom = 1, 29 → P-value: 3.342e-07



The scatter plot generated above displays less variation in LVO+ for large values of LOC compared to the non-logarithmic variables.

Model generated: LVO+ = 4.38 + 0.71(LOC), which displays a t-statistic of 6.57, where the resulting p-value < 0.05, indicating significant results. Therefore, there is enough evidence that LOC can be used as an explanatory variable to predict LVO+. The generated plot of residuals and QQ plot using LOC indicates the residuals are slightly curved and somewhat deviate from the normal distribution.

Using multiple linear regression model to predict LVO+ using LOC and LTRAP:

Residual Values Summary:

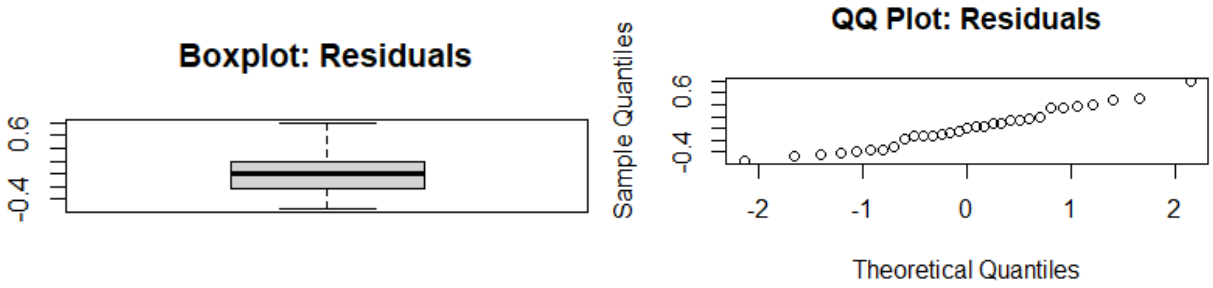
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.53385	-0.24194	-0.00337	0.18188	0.78501

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2592     0.3506  12.147 1.12e-12 ***
loc           0.4301     0.1680   2.560   0.0161 *
ltrap         0.4243     0.2054   2.066   0.0482 *
```

Residual Standard Error: 0.3394 with degrees of freedom = 28

Multiple R-squared: 0.6515, Adjusted R-squared: 0.6267

F-statistic: 26.18 with degrees of freedom = 2, 28 → P-value: 3.89e-07



Model: LVO+ = 4.26 + 0.43(LOC) + 0.42(LTRAP). The coefficient of LOC is statistically significant, since the t-value = 2.56, resulting in a p-value of 0.016, which is lower than the significance level of 0.05. In addition, the coefficient of LTRAP is statistically significant, with a t-value of 2.07 and a p-value of 0.048. Therefore, both LOC and LTRAP are good predictors of

LVO+, with LOC being more significant than LTRAP. This is consistent with the correlations found between the variables earlier in this exercise.

Using multiple linear regression model to predict LVO+ using LOC, LTRAP, and LVO-:

Model: $LVO+ = \beta_0 + \beta_1 (LOC) + \beta_2 (LTRAP) + \beta_3 (LVO-) + \varepsilon_i$, where ε_i are independent $N(0,\sigma)$ variables

Residual Values Summary:

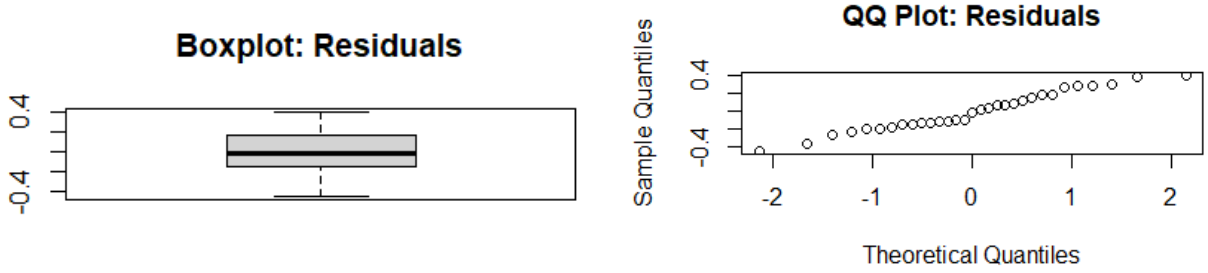
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.44029	-0.14718	-0.00694	0.16299	0.39917

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.87153    0.64015   1.361  0.18463
loc          0.39197    0.11535   3.398  0.00212 **
ltrap        0.02768    0.15697   0.176  0.86133
lvominus     0.67254    0.11779   5.710 4.56e-06 ***
```

Residual Standard Error: 0.2326 with degrees of freedom = 27

Multiple R-squared: 0.8421, Adjusted R-squared: 0.8246

F-statistic: 48.02 with degrees of freedom = 3, 27 → P-value: 5.906e-11



Since the results demonstrate p-values for the coefficients of LOC and LVO- below the significance level of 0.05, these variables represent strong explanatory variables in predicting LVO+. However, the p-value for LTRAP is 0.861, greater than the significance level, and does not produce significant results (a poor indicator for predicting LVO+). Clearly, the best explanatory variable in predicting LVO+ is LVO-.

Model			Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Explanatory Variable	LOC	Estimated Coefficient	0.7062	0.4301	0.39197
		Standard Error	0.1074	0.1680	0.11535
		t statistic	6.574	2.560	3.398
		p-value	3.34e-07	0.0161	0.00212
	LTRAP	Estimated Coefficient	N/A	0.4243	0.02768
		Standard Error		0.2054	0.15697
		t statistic		2.066	0.176
		p-value		0.0482	0.86133
	LVO-	Estimated Coefficient	N/A	N/A	0.67254
		Standard Error			0.11779
		t statistic			5.710
		p-value			4.56e-06

The coefficient for the LOC variable, which was used in all three regression models, decreased across each model. However, the p-value appears significant for all three models, where the values are less than the significance level of 0.05, indicating that the variable acts as a strong explanatory variable for predicting LVO+ in all models. The coefficient for the LTRAP variable, used in the second and third models, is significantly different between the two models, with a significant p-value in the second model but not the third. This indicates that the LTRAP variable acts as an explanatory variable in a model with LOC & LTRAP, but not with LOC & LTRAP & LVO-. Finally, the coefficient for the LVO- variable was only used in the third model, with a highly significant p-value, demonstrating that LVO- can be used to predict LVO+ when used in conjunction with LOC & LTRAP.

Model	Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Percent of Variation (adjusted R-Squared)	0.5846 → 58.46%	0.6267 → 62.67%	0.8246 → 82.46%
Estimate of σ (residual standard error)	0.358	0.3394	0.2326

The coefficient of determination, or the percent of variation, indicates the performance of the regression model. Therefore, the final multiple linear regression model exhibits the greatest R-Squared value, demonstrating that that model performs the best. Furthermore, the estimate of σ is the lowest for the final model, indicating the lowest residual standard error out of all three models.

Using multiple linear regression model to predict LVO+ using LOC and LVO-:

Residual Values Summary:

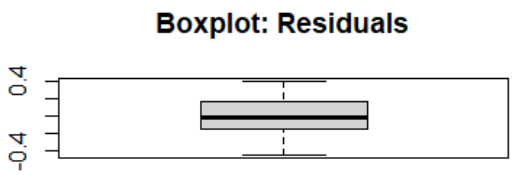
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.44129	-0.14493	-0.00965	0.16497	0.40145

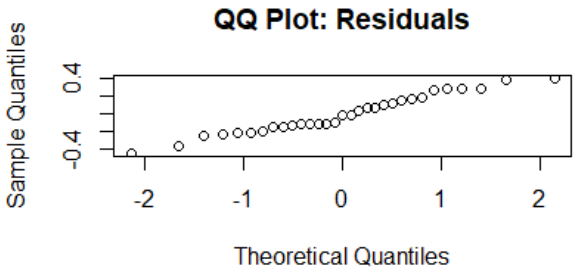
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.83180    0.58875   1.413    0.169
loc          0.40593    0.08242   4.925 3.40e-05 ***
lvominus     0.68173    0.10379   6.569 4.02e-07 ***
```

Residual Standard Error: 0.2286 with degrees of freedom = 28

Multiple R-squared: 0.842, Adjusted R-squared: 0.8307

F-statistic: 74.59 with degrees of freedom = 2, 28 → P-value: 6.061e-12





Since the results demonstrate p-values for the coefficients of LOC and LVO- below the significance level of 0.05, these variables represent strong explanatory variables in predicting LVO+. These results show even more significant p-values for the LOC and LVO- variables with the exclusion of the LTRAP variable compared to the previous model.

11.40

Using simple linear regression model to predict VO- using OC:

Residual Values Summary:

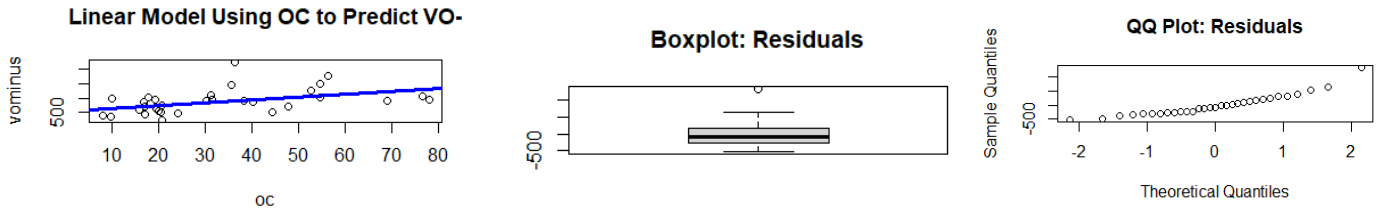
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-510.08	-276.24	-81.27	177.19	1317.22

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  557.818    139.151   4.009 0.000391 ***
oc             9.917      3.606   2.750 0.010161 *
```

Residual Standard Error: 387.4 with degrees of freedom = 29

Multiple R-squared: 0.2068, Adjusted R-squared: 0.1795

F-statistic: 7.561 with degrees of freedom = 1, 29 → P-value: 0.01016



The scatter plot generated above displays more variation in VO- for large values of OC. Model generated: VO- = 557.82 + 9.92(OC), which displays a t-statistic of 2.75, where the resulting p-value < 0.05, indicating significant results. Therefore, there is enough evidence that OC can be used as an explanatory variable to predict VO-. The generated plot of residuals and QQ plot using OC indicates the residuals are slightly curved and somewhat deviate from the normal distribution.

Using multiple linear regression model to predict VO- using OC and TRAP:

Residual Values Summary:

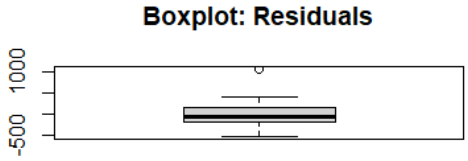
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-515.76	-194.82	-46.97	173.66	1068.57

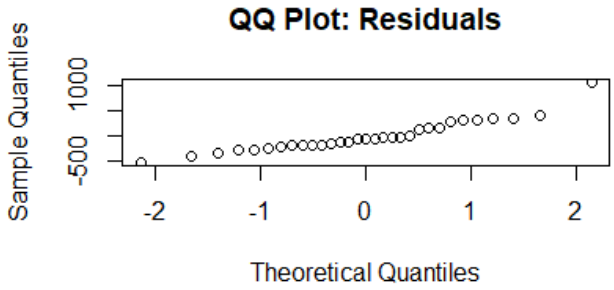
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  309.051    134.942   2.290  0.02974 *
oc            -1.868      4.418  -0.423  0.67567
trap          48.501     13.270   3.655  0.00105 **
```

Residual Standard Error: 324.4 with degrees of freedom = 28

Multiple R-squared: 0.463, Adjusted R-squared: 0.4247

F-statistic: 12.07 with degrees of freedom = 2, 28 → P-value: 0.0001658





Model: VO- = 309.05 - 1.87(OC) + 48.5(TRAP). The coefficient of OC is not statistically significant, since the t-value = -0.42, resulting in a p-value of 0.68, which is greater than the significance level of 0.05. However, the coefficient of TRAP is statistically significant, with a t-value of 3.66 and a p-value of 0.0011. Therefore, only TRAP is a good predictor of VO-, which is consistent with the correlations found between the variables.

Using multiple linear regression model to predict VO- using OC, TRAP, and VO+:

Model: VO- = $\beta_0 + \beta_1$ (OC) + β_2 (TRAP) + β_3 (VO+) + ε_i, where ε_i are independent N(0,σ) variables

Residual Values Summary:

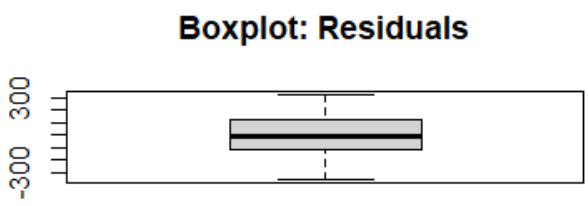
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-346.99	-111.42	-4.38	118.33	317.70

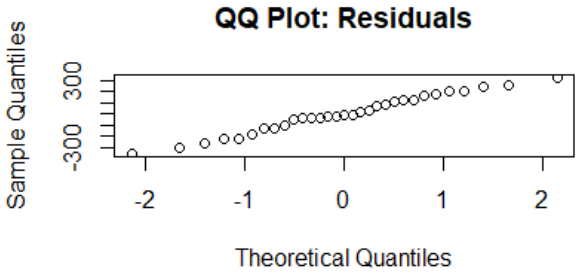
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 267.26110   74.71782   3.577  0.00134 **
oc           -6.51323    2.50744  -2.598  0.01502 *
trap          9.48453    8.78782   1.079  0.29001
voplus        0.72420    0.08999   8.048  1.2e-08 ***
```

Residual Standard Error: 179.2 with degrees of freedom = 27

Multiple R-squared: 0.842, Adjusted R-squared: 0.8245

F-statistic: 47.97 with degrees of freedom = 3, 27 → P-value: 5.974e-11





Since the results demonstrate p-values for the coefficients of OC and VO+ below the significance level of 0.05, these variables represent strong explanatory variables in predicting VO-. However, the p-value for TRAP is 0.861, greater than the significance level, and does not produce significant results. Clearly, the best explanatory variable in predicting VO- is VO+.

Model			Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Explanatory Variable	OC	Estimated Coefficient	9.917	-1.868	-6.51323
		Standard Error	3.606	4.418	2.50744
		t statistic	2.750	-0.423	-2.598
		p-value	0.010161	0.67567	0.01502
	TRAP	Estimated Coefficient	N/A	48.501	9.48453
		Standard Error		13.270	8.78782
		t statistic		3.655	1.079
		p-value		0.00105	0.29001
	VO+	Estimated Coefficient	N/A	N/A	0.72420
		Standard Error			0.08999

		t statistic			8.048
		p-value			1.2e-08

The coefficient for the OC variable, which was used in all three regression models, decreased across each model. However, the p-value appears significant for the first and third models, where the values are less than the significance level of 0.05, indicating that the variable acts as a strong explanatory variable for predicting VO- in two of the three models. The coefficient for the TRAP variable, used in the second and third models, is significantly different between the two models, with a significant p-value in the second model but not the third. This indicates that the TRAP variable acts as an explanatory variable in a model with OC and TRAP, but not with OC and TRAP and VO+. Finally, the coefficient for the VO+ variable was only used in the third model, with a highly significant p-value, demonstrating that VO+ can be used to predict VO- when used in conjunction with OC and TRAP.

Model	Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Percent of Variation (adjusted R-Squared)	0.1795 → 17.95%	0.4247 → 42.47%	0.8245 → 82.45%
Estimate of σ (residual standard error)	387.4	324.4	179.2

The coefficient of determination, or the percent of variation, indicates the performance of the regression model. Therefore, the final multiple linear regression model exhibits the greatest R-Squared value, demonstrating that that model performs the best. Furthermore, the estimate of σ is the lowest for the final model, indicating the lowest residual standard error out of all three models.

Using multiple linear regression model to predict VO- using OC and VO+:

Residual Values Summary:

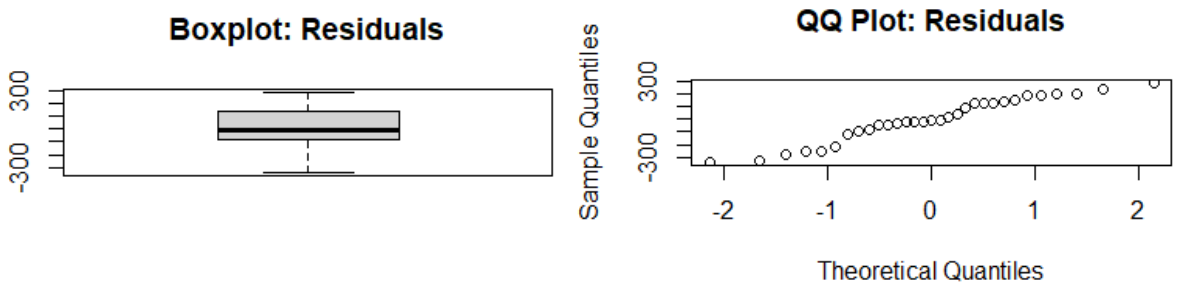
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-334.86	-84.97	-6.61	135.47	285.01

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 298.01211   69.27509   4.302 0.000186 ***
oc           -5.25375    2.22586  -2.360 0.025459 *
voplus        0.77778    0.07527  10.333 4.65e-11 ***
```

Residual Standard Error: 179.7 with degrees of freedom = 28

Multiple R-squared: 0.8352, Adjusted R-squared: 0.8234

F-statistic: 70.95 with degrees of freedom = 2, 28 → P-value: 1.09e-11



Since the results demonstrate p-values for the coefficients of OC and VO+ below the significance level of 0.05, these variables represent strong explanatory variables in predicting VO-. These results show even more significant p-values for the OC and VO+ variables with the exclusion of the TRAP variable compared to the previous model.

11.41

Using simple linear regression model to predict LVO- using LOC:

Residual Values Summary:

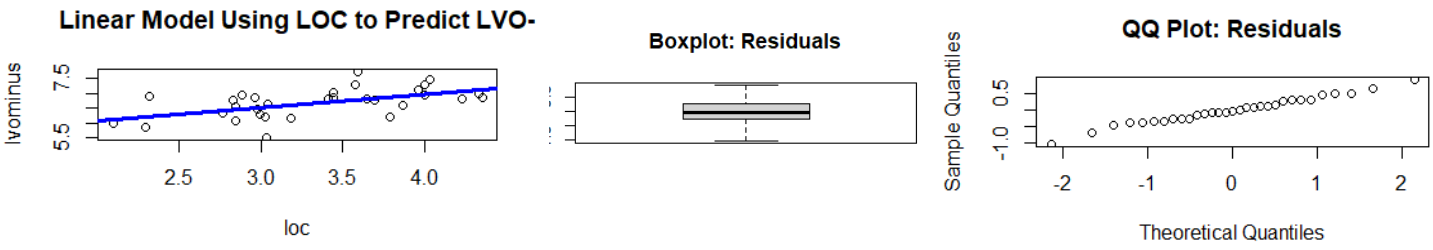
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-1.0111	-0.2662	-0.0472	0.2938	0.9174

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.2115     0.4161  12.524 3.19e-13 ***
loc           0.4404     0.1227   3.589   0.0012 **
```

Residual Standard Error: 0.409 with degrees of freedom = 29

Multiple R-squared: 0.3076, Adjusted R-squared: 0.2837

F-statistic: 12.88 with degrees of freedom = 1, 29 → P-value: 0.001205



The scatter plot generated above displays less variation in LVO- for large values of LOC compared to the non-logarithmic variables.

Model generated: LVO- = 5.21 + 0.44(LOC), which displays a t-statistic of 3.59, where the resulting p-value < 0.05, indicating significant results. Therefore, there is enough evidence that LOC can be used as an explanatory variable to predict LVO-. The generated plot of residuals and QQ plot using LOC indicates the residuals are mostly linear and only slightly deviate from the normal distribution.

Using multiple linear regression model to predict LVO- using LOC and LTRAP:

Residual Values Summary:

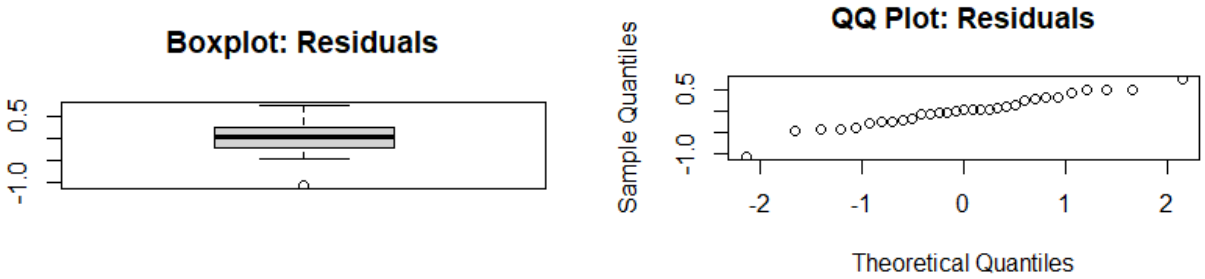
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-1.04755	-0.23063	0.02051	0.24895	0.73126

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.03718    0.38559  13.063 1.96e-13 ***
loc          0.05675    0.18476   0.307   0.7610
ltrap        0.58969    0.22584   2.611   0.0143 *
```

Residual Standard Error: 0.3733 with degrees of freedom = 28

Multiple R-squared: 0.4432, Adjusted R-squared: 0.4034

F-statistic: 11.14 with degrees of freedom = 2, 28 → P-value: 0.0002755



Model: LVO- = 5.04 + 0.057(LOC) + 0.59(LTRAP). The coefficient of LOC is not statistically significant, since the t-value = 0.31, resulting in a p-value of 0.76, which is greater than the significance level of 0.05. However, the coefficient of LTRAP is statistically significant, with a t-value of 2.61 and a p-value of 0.014. Therefore, only LTRAP is a good predictor of LVO-, which is consistent with the correlations found between the variables.

Using multiple linear regression model to predict LVO- using LOC, LTRAP, and LVO+:

Model: LVO- = $\beta_0 + \beta_1$ (LOC) + β_2 (LTRAP) + β_3 (LVO+) + ε_i, where ε_i are independent N(0,σ) variables

Residual Values Summary:

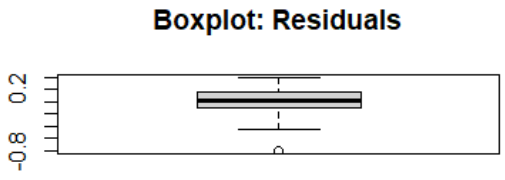
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.79211	-0.09676	0.01604	0.15190	0.38693

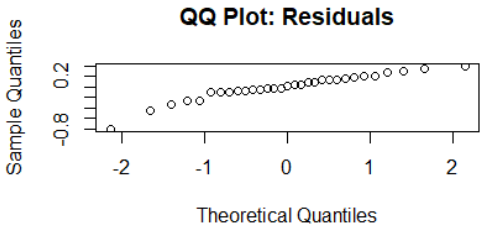
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.5731     0.6618   2.377   0.0248 *
loc          -0.2931     0.1407  -2.083   0.0468 *
ltrap         0.2446     0.1662   1.472   0.1526
lvoplus       0.8133     0.1424   5.710 4.56e-06 ***
```

Residual Standard Error: 0.2558 with degrees of freedom = 27

Multiple R-squared: 0.7478, Adjusted R-squared: 0.7197

F-statistic: 26.68 with degrees of freedom = 3, 27 → P-value: 3.133e-08





Since the results demonstrate p-values for the coefficients of LOC and LVO+ below the significance level of 0.05, these variables represent strong explanatory variables in predicting LVO-. However, the p-value for LTRAP is 0.153, greater than the significance level, and does

not produce significant results (a poor indicator for predicting LVO-). Clearly, the best explanatory variable in predicting LVO- is LVO+.

Model			Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Explanatory Variable	LOC	Estimated Coefficient	0.4404	0.05675	-0.2931
		Standard Error	0.1227	0.18476	0.1407
		t statistic	3.589	0.307	-2.083
		p-value	0.0012	0.7610	0.0468
	LTRAP	Estimated Coefficient	N/A	0.58969	0.2446
		Standard Error		0.22584	0.1662
		t statistic		2.611	1.472
		p-value		0.0143	0.1526
	LVO+	Estimated Coefficient	N/A	N/A	0.8133
		Standard Error			0.1424
		t statistic			5.710
		p-value			4.56e-06

The coefficient for the LOC variable, which was used in all three regression models, decreased across each model. However, the p-value appears significant for the first and third models, where the values are less than the significance level of 0.05, indicating that the variable acts as a strong explanatory variable for predicting LVO- in two of the three models. The coefficient for the LTRAP variable, used in the second and third models, is significantly different between the two models, with a significant p-value in the second model but not the third. This indicates that the LTRAP variable acts as an explanatory variable in a model with LOC and LTRAP, but not with LOC and LTRAP and LVO+. Finally, the coefficient for the LVO+ variable was only used in the third model, with a highly significant p-value, demonstrating that LVO+ can be used to predict LVO- when used in conjunction with LOC and LTRAP.

Model	Linear Regression	Multiple Linear Regression 1	Multiple Linear Regression 2
Percent of Variation (adjusted R-Squared)	0.2837 → 28.37%	0.4034 → 40.34%	0.7197 → 71.97%
Estimate of σ (residual standard error)	0.409	0.3733	0.2558

The coefficient of determination, or the percent of variation, indicates the performance of the regression model. Therefore, the final multiple linear regression model exhibits the greatest R-Squared value, demonstrating that that model performs the best. Furthermore, the estimate of σ is the lowest for the final model, indicating the lowest residual standard error out of all three models.

Using multiple linear regression model to predict LVO- using LOC and LVO+:

Residual Values Summary:

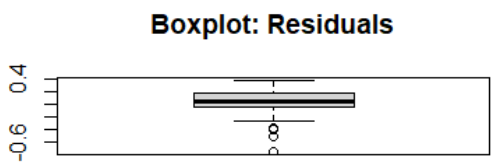
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-0.75505	-0.04466	0.03346	0.16848	0.36414

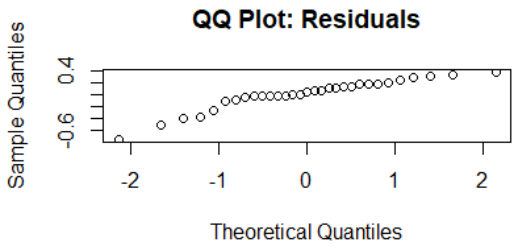
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.3110     0.6505   2.015   0.0536 .
loc          -0.1878     0.1236  -1.519   0.1400
lvoplus       0.8896     0.1354   6.569 4.02e-07 ***
```

Residual Standard Error: 0.2611 with degrees of freedom = 28

Multiple R-squared: 0.7275, Adjusted R-squared: 0.708

F-statistic: 37.38 with degrees of freedom = 2, 28 → P-value: 1.245e-08





Since the p-value for the coefficient of LVO+ is below the significance level of 0.05, this variable represents a strong explanatory variable in predicting VO-. However, in this model, LOC does not represent a strong explanatory variable, with a p-value greater than 0.05.