Final Project

Colleen Que, Ron Dumalagan
12/16/21
I pledge my honor that I have abided by the Stevens Honor System.

Summary of Conclusion: This project specifically focuses on the analysis of healthy bone renewal in a study of 31 healthy women between the ages of 11 to 32. The dataset being used contains the following data from each of the 31 participants: VO+ and VO- (which measure bone formation and bone resorption respectively), OC, and TRAP. In addition, the dataset contains LVO+, LVO-, LOC, and LTRAP, which are the logarithms of each of the measured variables. To aid with calculations and forming both numerical and graphical representations of our data, we utilized *RStudio* (version 1.41717), an IDE for the *R* programming language.

Description of the data: Healthy bones are continually renewed through two processes: bone formation (new bone being built), and bone resorption (old bone being removed). These processes are measured in order to determine if a person is experiencing bone loss, which could potentially be a result of disease, aging, or space travel. In order to measure bone formation and bone resorption, the biochemical markers osteocalcin (OC) and tartrate-resistant acid phosphatase (TRAP) are used, since both markers are positively associated with bone formation and bone resorption respectively. OC is measured in milligrams of OC per milliliter of blood (mg/ml), whereas TRAP is measured in units per liter (U/l).

Methods (model) of Statistical Analysis: We first created numerical and graphical summaries of VO+, VO-, OC, and TRAP, in order to analyze the distributions of each variable. The numerical summaries consisted of the minimum, maximum, median, mean, 1st quartile, and 3rd quartile values of each variable. The graphical summaries consisted of stem-and-plot graphs of each variable. Based on our summaries, we were able to determine that the distributions of all of these variables are slightly skewed to the right. After creating scatter plots and calculating the correlation coefficients between each of the variables, we were then able to determine that all of the relationships between each pair of variables are positively associated. The relationship between VO+ and VO- has the strongest association, whereas the relationship between VO- and OC has the weakest association.

We then created a simple linear regression model in order to predict VO+, with OC as the explanatory variable. Through the scatter plots we created, we noticed that there appears to be more variation in VO+ values as OC increases, and this trend continued to be true in our regression model. Since the processes of bone formation and bone resorption are highly related, we continued by creating a multiple regression model that used both OC and TRAP as explanatory variables in order to predict VO+. This model helped us confirm that, despite being the biochemical marker for bone resorption, TRAP is a better predictor of bone formation than OC is, as it has a higher correlation with VO+ than OC does, lower p-value than OC , and higher coefficient than OC in the model.

Next, we explored another multiple regression model, this time using OC, TRAP, and VO- as explanatory variables in order to predict VO+. In this model, we found that TRAP was not a significant variable, suggesting that we run another model without TRAP, only using OC and VO-. Comparing the model with TRAP to the model without TRAP, both models provided similar results, as the model without TRAP only had a slightly lower R^2 and standard error. As it is a simpler model, we conclude that the model without TRAP, only using OC and VO- as explanatory variables, is the best model for predicting VO+.

Due to the skewness of the distributions of VO+, VO−, OC, and TRAP that we discovered in our earlier summaries of the variables, we then proceeded to redo the entire procedure, but this time using the logarithms of each variable instead. It is common to work with logarithms rather than the measured values, in the case that the original values are skewed. As opposed to the original data, the logarithmic values appear to have more symmetrical distributions. Based on the scatterplots, all of the relationships between each pair of logarithm variables are positively associated. The relationship between LVO+ and LVO- has the strongest association, whereas the relationship between LOC and LVO- has the weakest association. We came to a similar conclusion with the regression models of the logarithm variables, that LTRAP is not significant. As the R^2 value of the model without LTRAP is nearly identical to the model with LTRAP , we can say that this model (just using LOC and LVO-) is the best predictor of LVO+.

Finally, we explored another multiple regression model, this time using OC, TRAP, and VO+ as explanatory variables in order to predict VO-. In this model, we found that TRAP was not a significant variable, suggesting that we run another model without TRAP, only using OC and VO+. Variation for estimation without using trap is slightly lower than estimation with trap (0.835< 0.842), and standard error of estimation without using trap is slightly larger than the one in part (b) (179.7>179.2), which are similar results. As it is a simpler model, we conclude that the model without TRAP, only using OC and VO+ as explanatory variables, is the best model for predicting VO-. Similar to predicting VO+, we proceed to use logarithms to redo the procedure due to skewness of distributions.

As opposed to the original data, the logarithmic values appear to have more symmetrical distributions. Based on the scatterplots, all of the relationships between each pair of logarithm variables are positively associated. We came to a similar conclusion with the regression models of the logarithm variables, that LTRAP is not significant. As the R^2 value of the model without LTRAP is nearly identical to the model with LTRAP , we can say that this model (just using LOC and LVO+) is the best predictor of LVO-.

Version of the Software: RStudio 2021.09.1 Build 372

Conclusion: Predictions for VO- and VO+ are very similar. In both cases, we have concluded that using a logarithmic model without the variable trap is the best predictor for the goal variable. In this project, we learned that running multiple regressions can help us determine what model is the best for predicting a variable using R-studio. The graphs and output datas from summary and anova make the relationships between data very clear and is beneficial for the future if there are any other studies that involve multiple variables influencing each other.

Pros and Cons: A logarithmic model is preferred over a normal one due to the skewness of the data. This model is accurate because we only neglected the variable with a high p-value. However, in order to obtain this model, we had to run multiple regressions which takes up time. The neglected value ltrap might have some influence on the final result, which then limits the precision of the model.

11.36 Bone formation and resorption.

(a)

Based on the numerical and graphical summaries, we can conclude that the distributions of all four variables are slightly skewed to the right.

```
       voplus                  vominus                    oc                     trap
Min.    : 285.0        Min.     : 254.0        Min.     : 8.10        Min.     : 3.30
1st Qu.: 542.5        1st Qu.: 554.0        1st Qu.:18.60        1st Qu.: 8.90
Median : 870.0        Median : 903.0        Median :30.20        Median :10.30
Mean    : 985.8        Mean     : 889.2        Mean     :33.42        Mean     :13.25
3rd Qu.:1188.5        3rd Qu.:1023.0        3rd Qu.:46.05        3rd Qu.:18.80
Max.    :2545.0        Max.     :2236.0        Max.     :77.90        Max.     :28.80
```

Stemplots

```
VO+:                  VO-:                  OC:                   Trap:

 2 | 99                2 | 55                0 | 8                 0 | 3
 4 | 0588117           4 | 1589047           1 | 00677789          0 | 667888999
 6 | 23995             6 | 3027              2 | 001114            1 | 000000114
 8 | 478079            8 | 670223568         3 | 011668            1 | 55599999
10 | 303              10 | 2372              4 | 048               2 | 4
12 | 50               12 | 5                 5 | 3556              2 | 569
14 | 6                14 | 59                6 | 9
16 | 18               16 | 6                 7 | 78
18 |                  18 |
20 |                  20 |
22 | 24               22 | 4
24 | 5
```

(b)

Based on the scatterplots, all of the relationships between each pair of variables are positively associated. The relationship between VO+ and VO- has the strongest association with a correlation coefficient of 0.8957707, whereas the relationship between VO- and OC has the weakest association with a correlation coefficient of 0.4547603.

Correlation coefficients between each pair of variables:

```
> cor(voplus, vominus)
[1] 0.8957707
> cor(voplus, oc)
[1] 0.659614
> cor(voplus, trap)
[1] 0.7648649
> cor(vominus, oc)
[1] 0.4547603
> cor(vominus, trap)
[1] 0.6779267
> cor(oc, trap)
[1] 0.7298519
```


VO+ vs VO-

VO+ vs OC
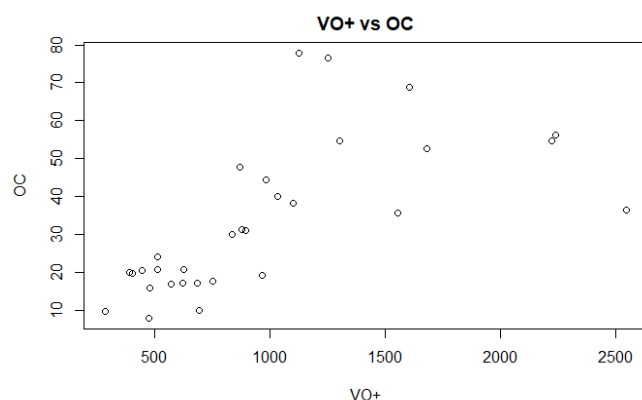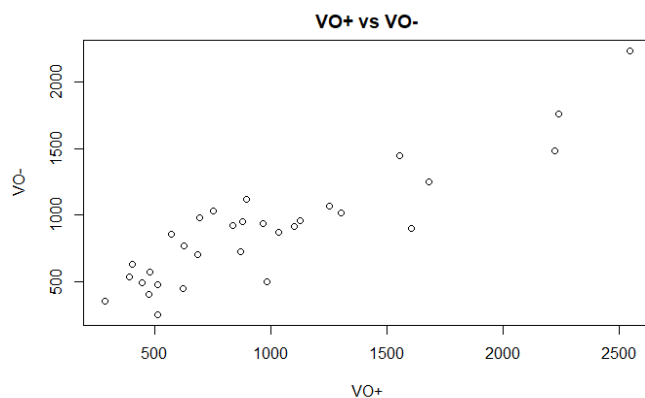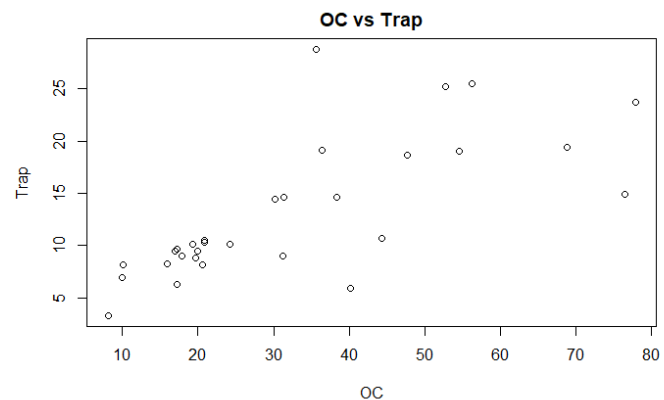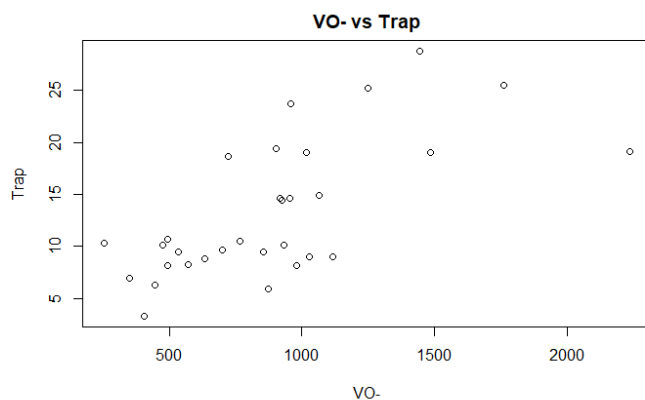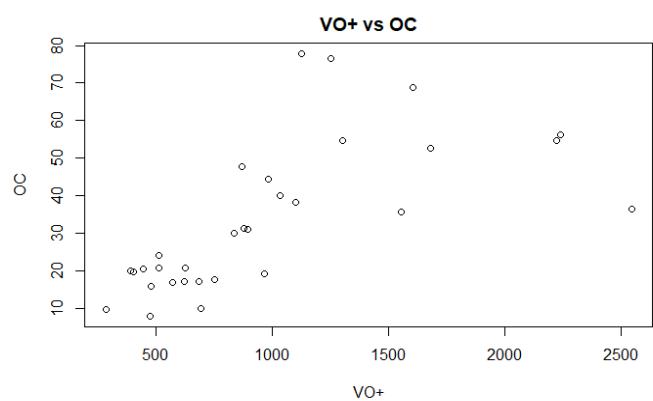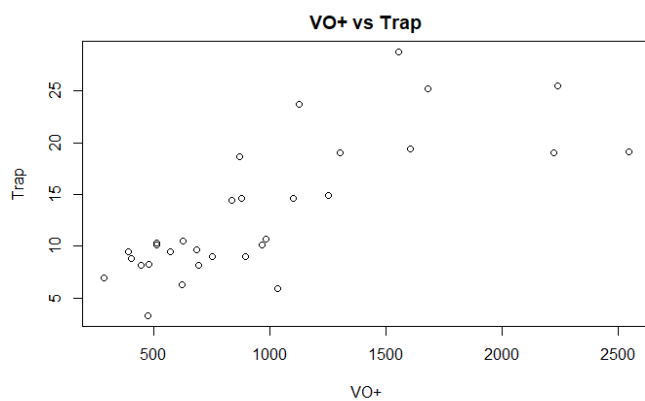
11.37 Predicting bone formation.

(a)

According to the data from the previous question, there appears to be more variation in VO+ values as OC increases. This pattern continues to be true when running this regression model. The plot of the residuals vs. OC begins to curve downward towards the right end. As shown in the Q-Q plot, the residuals appear to be skewed to the right.

$$\widehat{VO+} = 334.034 + 19.505\,OC$$

```
> summary(linreg)

Call:
lm(formula = voplus ~ oc)

Residuals:
    Min      1Q  Median      3Q     Max
-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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 443.3 on 29 degrees of freedom
Multiple R-squared:  0.4351,    Adjusted R-squared:  0.4156
F-statistic: 22.34 on 1 and 29 DF,  p-value: 5.429e-05
```
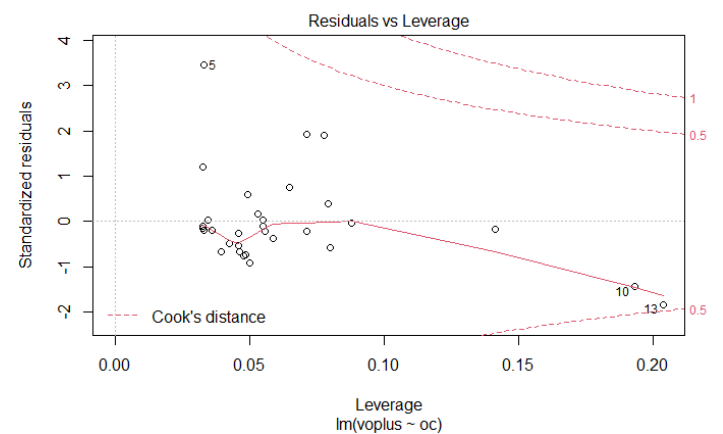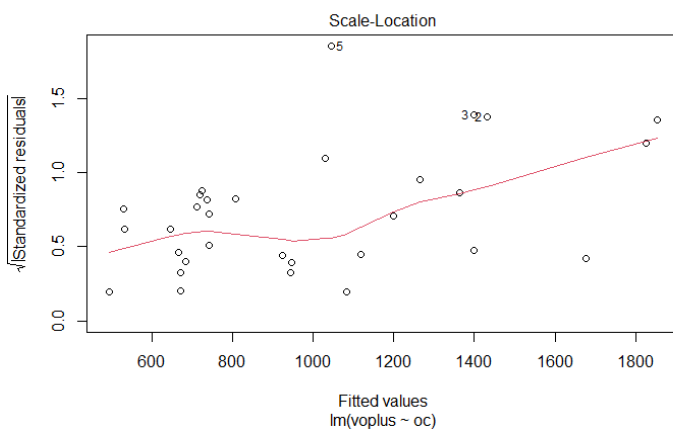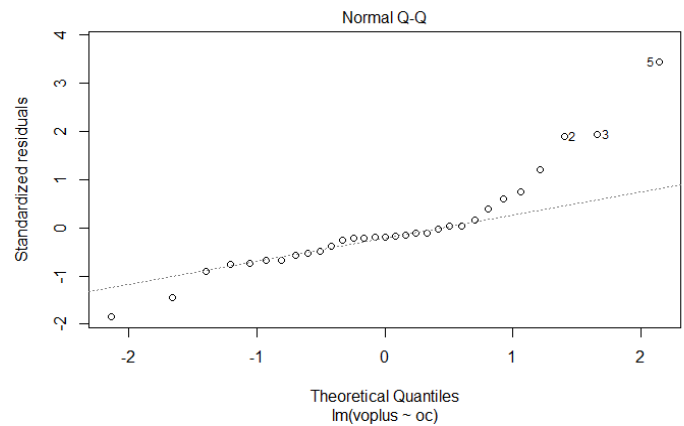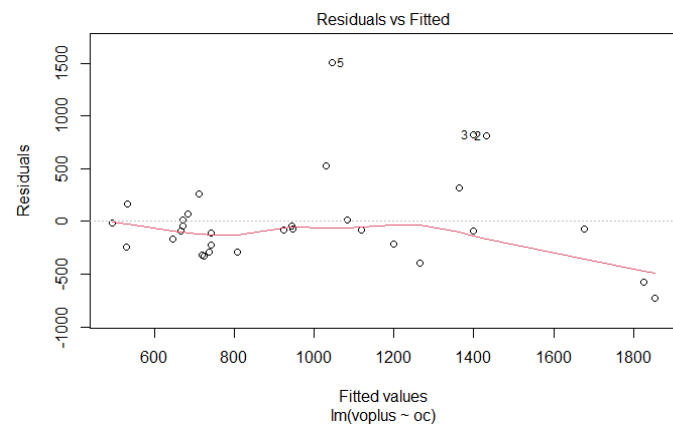
Residual standard error: 443.3
Multiple R^2 = 0.4351 Adjusted R^2: 0.4156

(b)

This view is consistent with the pattern of relationships described in the previous exercise, as TRAP has a stronger correlation with VO+ than OC does. The coefficient of TRAP is also much greater than the coefficient of OC, meaning it is more significant when both are present in the model.

$$\widehat{VO+} = 57.704 \ + \ 6.415 \, OC \ + \ 53.874 \, TRAP$$

```
> summary(linreg)

Call:
lm(formula = voplus ~ oc + trap)

Residuals:
    Min      1Q  Median      3Q     Max
-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 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 376.3 on 28 degrees of freedom
Multiple R-squared:  0.607,     Adjusted R-squared:  0.5789
F-statistic: 21.62 on 2 and 28 DF,  p-value: 2.096e-06
```

$R^2 = 0.607$, Residual standard error = 376.3

11.38 More on predicting bone formation.

(a)

VO+ = β_0 + β_1OC + β_2TRAP + β_3 VO- + ε_i i = 1, ... ,n
Assumption: ε are independent N(0,σ) variables.

(b)

Due to larger p-value (0.52797), trap is not significant.

Estimated equation: VO+ = -243.4877 + 0.9746 VO- + 6.6071 trap + 8.2349 OC + ε_i
Residual standard error = 207.8 on 27 degrees of freedom
Multiple R^2 = 0.8844, Adjusted R^2 = 0.8715
F statistic = 68.84 , p value = 9.031e-13

```
> summary(lm1)

Call:
lm(formula = voplus ~ vominus + trap + oc, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-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 *
vominus        0.9746     0.1211   8.048  1.2e-08 ***
trap           6.6071    10.3340   0.639  0.52797
oc             8.2349     2.8397   2.900  0.00733 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 207.8 on 27 degrees of freedom
Multiple R-squared:  0.8844,    Adjusted R-squared:  0.8715
F-statistic: 68.84 on 3 and 27 DF,  p-value: 9.031e-13

> anova(lm1)
Analysis of Variance Table

Response: voplus
          Df  Sum Sq Mean Sq  F value    Pr(>F)
vominus    1 8093909 8093909 187.3768 1.153e-13 ***
trap       1  463595  463595  10.7324  0.002891 **
oc         1  363268  363268   8.4098  0.007332 **
Residuals 27 1166289   43196
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c)

Model	Variable Coefficients	Variable Standard Error	Variable t-statistic	Variable p-value	R^2	s
OC	OC: 19.505	OC: 4.13	OC: 4.73	OC: p < 0.05	0.4351	443.3
OC, TRAP	OC: 6.415 TRAP: 53.874	OC: 5.13 TRAP: 15.39	OC: 1.25 TRAP: 3.50	OC: 0.22 TRAP: 0.002	0.607	376.3
OC, TRAP, VO-	OC: 8.24 TRAP: 6.61 VO-: 0.9746	OC: 2.84 TRAP: 10.33 VO-: 0.12	OC: 2.9 TRAP: 0.64 VO-: 8.05	OC: 0.007 TRAP: 0.53 VO-: p < 0.05	0.8844	207.8

(d)

- OC:
    - Variation: Multiple R^2 = 0.4351 = 43.51%
    - Residual standard error=σ= 443.3
- OC, TRAP:
    - Variation: Multiple R^2 = 0.607 = 60.7%
    - Residual standard error = σ= 376.3
- OC,TRAP, VO-:
    - Variation: Multiple R^2 = 0.8844= 88.44%
    - σ = 207.8

- Summary: Percent of variation in VO+ explained by OC is the lowest, but the standard error is the highest. Percent of variation in VO+ explained by OC, Trap, and VO- is the highest, but the standard error is the lowest.

(e)

The other model is a model without TRAP, since TRAP has the highest p-value.

```
Call:
lm(formula = voplus ~ vominus + oc, data = project)

Residuals:
    Min      1Q  Median      3Q     Max
-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 *
vominus        1.01857    0.09858  10.333 4.65e-11 ***
oc             9.40388    2.14964   4.375 0.000153 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 205.6 on 28 degrees of freedom
Multiple R-squared:  0.8826,     Adjusted R-squared:  0.8742
F-statistic: 105.3 on 2 and 28 DF,  p-value: 9.418e-14
```

```
Analysis of Variance Table

Response: voplus
          Df  Sum Sq Mean Sq F value     Pr(>F)
vominus    1 8093909 8093909 191.419 4.835e-14 ***
oc         1  809205  809205  19.137 0.0001528 ***
Residuals 28 1183947   42284
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result:
Residual standard error: 205.6 on 28 degrees of freedom
Multiple R-squared: 0.8826,   Adjusted R-squared: 0.8742

From part (b): OC,TRAP, VO-:
- Variation: Multiple R^2 = 0.8844= 88.44%
- σ = 207.8

Variation for estimation without using trap is slightly lower than estimation with trap (0.8826< 0.8844), and standard error of estimation without using trap is also slightly lower than the one in part (b) (205.6<207.8). The values are very similar.

11.39 Predicting bone formation using transformed variables.

Because the distributions of VO+, VO−, OC, and TRAP tend to be skewed, it is common to work with logarithms rather than the measured values. Using the questions in the previous three exercises as a guide, analyze the log data.

As opposed to the original data, the logarithmic values appear to have more symmetrical distributions.

```
        loc                ltrap              lvoplus             lvominus
Min.    :2.092     Min.    :1.194     Min.     :5.652     Min.    :5.537
1st Qu.:2.922      1st Qu.:2.186      1st Qu.:6.295       1st Qu.:6.316
Median :3.408      Median :2.332      Median :6.768       Median :6.806
Mean    :3.338     Mean    :2.467     Mean     :6.742     Mean    :6.682
3rd Qu.:3.829      3rd Qu.:2.933      3rd Qu.:7.079       3rd Qu.:6.931
Max.    :4.355     Max.    :3.360     Max.     :7.842     Max.    :7.712
```

Stemplots

LOC:	LTRAP:	LVO+:	LVO-:
20 | 9	10 | 9	56 | 5	54 | 4
22 | 91	12 |	58 | 7	56 |
24 |	14 |	60 | 0077	58 | 6
26 | 7	16 | 8	62 | 445	60 | 107
28 | 344868	18 | 43	64 | 4434	62 | 0185
30 | 03349	20 | 9907	66 | 2378	64 | 558
32 |	22 | 0055611357	68 | 0894	66 | 457
34 | 14589	24 |	70 | 0337	68 | 1234679347
36 | 599	26 | 7880	72 | 58	70 | 23
38 | 67	28 | 24457	74 | 3	72 | 80
40 | 003	30 | 7	76 | 11	74 | 7
42 | 346	32 | 346	78 | 4	76 | 1

Based on the scatterplots, all of the relationships between each pair of logarithm variables are positively associated. The relationship between LVO+ and LVO- has the strongest association with a correlation coefficient of 0.8957707, whereas the relationship between LOC and LVO- has the weakest association with a correlation coefficient of 0.554607.

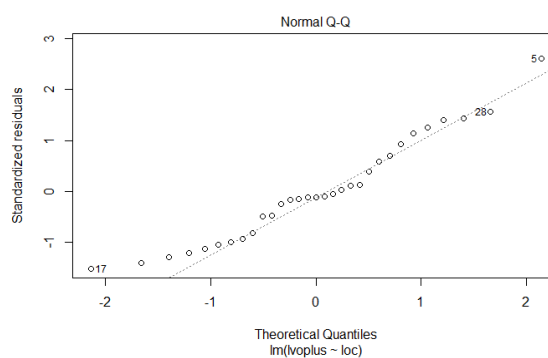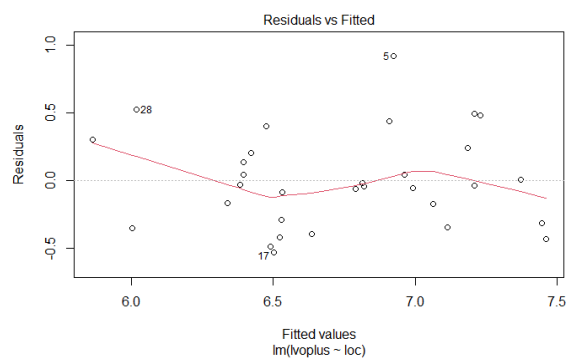Correlation coefficients between each pair of variables:
```
> cor(loc, ltrap)
[1] 0.7953528
> cor(loc, lvoplus)
[1] 0.7735853
> cor(loc, lvominus)
[1] 0.554607
> cor(ltrap, lvoplus)
[1] 0.7549684
> cor(ltrap, lvominus)
[1] 0.6643005
> cor(lvoplus, lvominus)
[1] 0.8396741
```
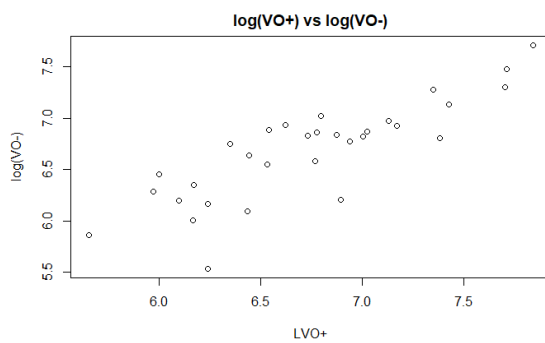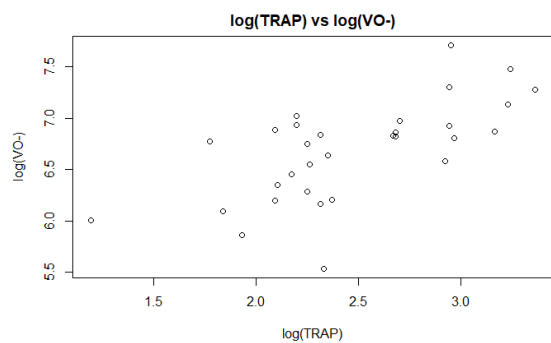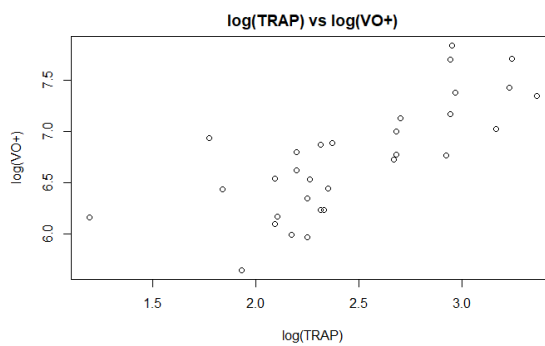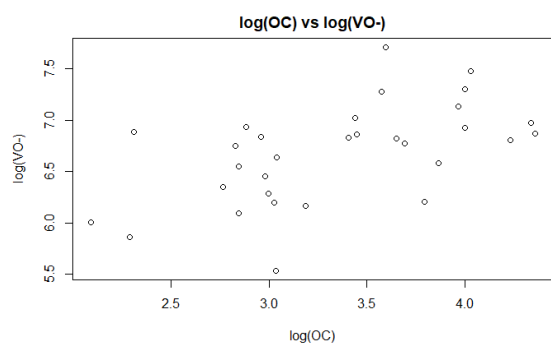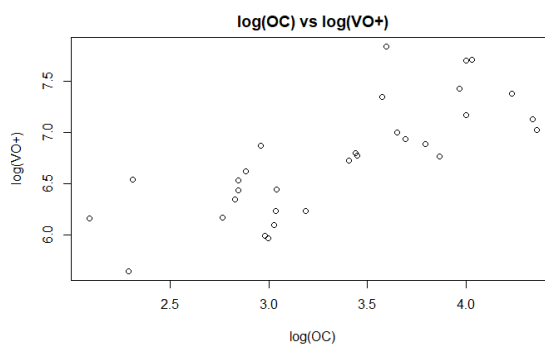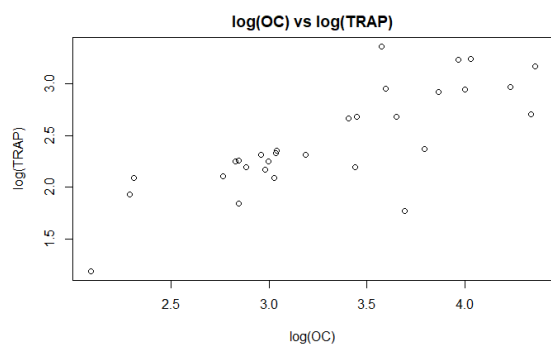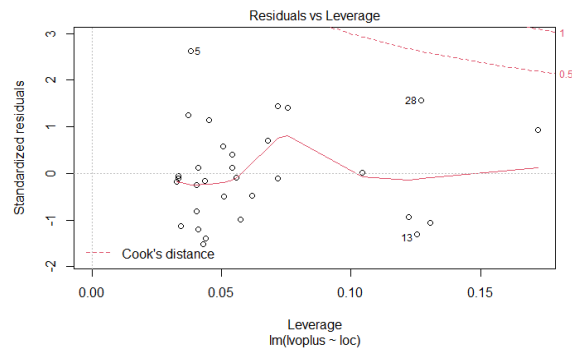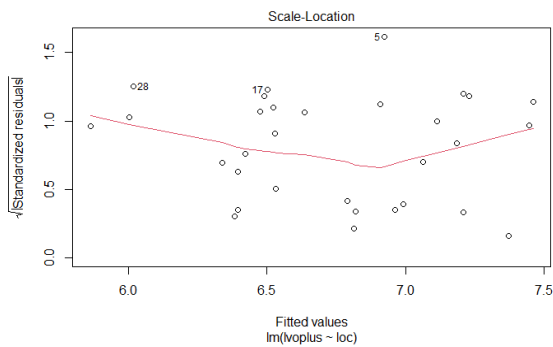
Regression models:

$$\widehat{LVO+} = 4.3846 + 0.7062\, LOC$$

$$\widehat{LVO+} = 4.2592 + 0.4301 \, LOC + 0.4243 \, LTRAP$$

$$\widehat{LVO+} = 0.87153 + 0.39197 \, LOC + 0.02768 \, LTRAP + 0.67254 \, LVO-$$

Model	Variable Coefficients	Variable Standard Error	Variable t-statistic	Variable p-value	R^2	s
LOC	LOC: 0.706	LOC: 0.107	LOC: 6.574	LOC: p < 0.05	0.598	0.358
LOC, LTRAP	LOC: 0.430 LTRAP: 0.424	LOC: 0.168 LTRAP: 0.205	LOC: 2.56 LTRAP: 2.066	LOC: 0.016 LTRAP: 0.048	0.652	0.339
LOC, LTRAP, LVO-	LOC: 0.392 LTRAP: 0.027 LVO-: 0.672	LOC: 0.115 LTRAP: 0.156 LVO-: 0.118	LOC: 3.398 LTRAP: 0.176 LVO-: 5.710	LOC: 0.002 LTRAP: 0.861 LVO-: p < 0.05	0.842	0.233

Due to the significantly higher p-value of LTRAP (p = 0.861), we can make a similar conclusion to the previous exercise that LTRAP is not significant.

Model with LTRAP removed:
$$\widehat{LVO+} = 0.83180 + 0.40593 \, LOC + 0.68173 \, LVO-$$

```
Call:
lm(formula = lvoplus ~ loc + lvominus)

Residuals:
     Min       1Q    Median       3Q       Max
-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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2286 on 28 degrees of freedom
Multiple R-squared:  0.842,      Adjusted R-squared:  0.8307
F-statistic: 74.59 on 2 and 28 DF,  p-value: 6.061e-12
```

As the R^2 value of the model with LTRAP removed is nearly identical to the model with LOC, LTRAP, and LVO-, we can say that this model (just using LOC and LVO-) is the best predictor of LVO+.

11.40 Predicting bone resorption.

$$\widehat{VO-} = 557.818 + 9.917\,OC$$

Residual standard error: 387.4 on 29 degrees of freedom

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

F-statistic: 7.561 on 1 and 29 DF, p-value: 0.01016

```
    Min      1Q  Median      3Q     Max
-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 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 387.4 on 29 degrees of freedom
Multiple R-squared:  0.2068,     Adjusted R-squared:  0.1795
F-statistic: 7.561 on 1 and 29 DF,  p-value: 0.01016

Analysis of Variance Table

Response: vominus
          Df  Sum Sq Mean Sq F value  Pr(>F)
oc         1 1134474 1134474  7.5611 0.01016 *
Residuals 29 4351193  150041
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual Plots:

The plot of the residuals vs. OC begins to curve upwards towards both left and right ends. The residuals appear to be skewed to the right.

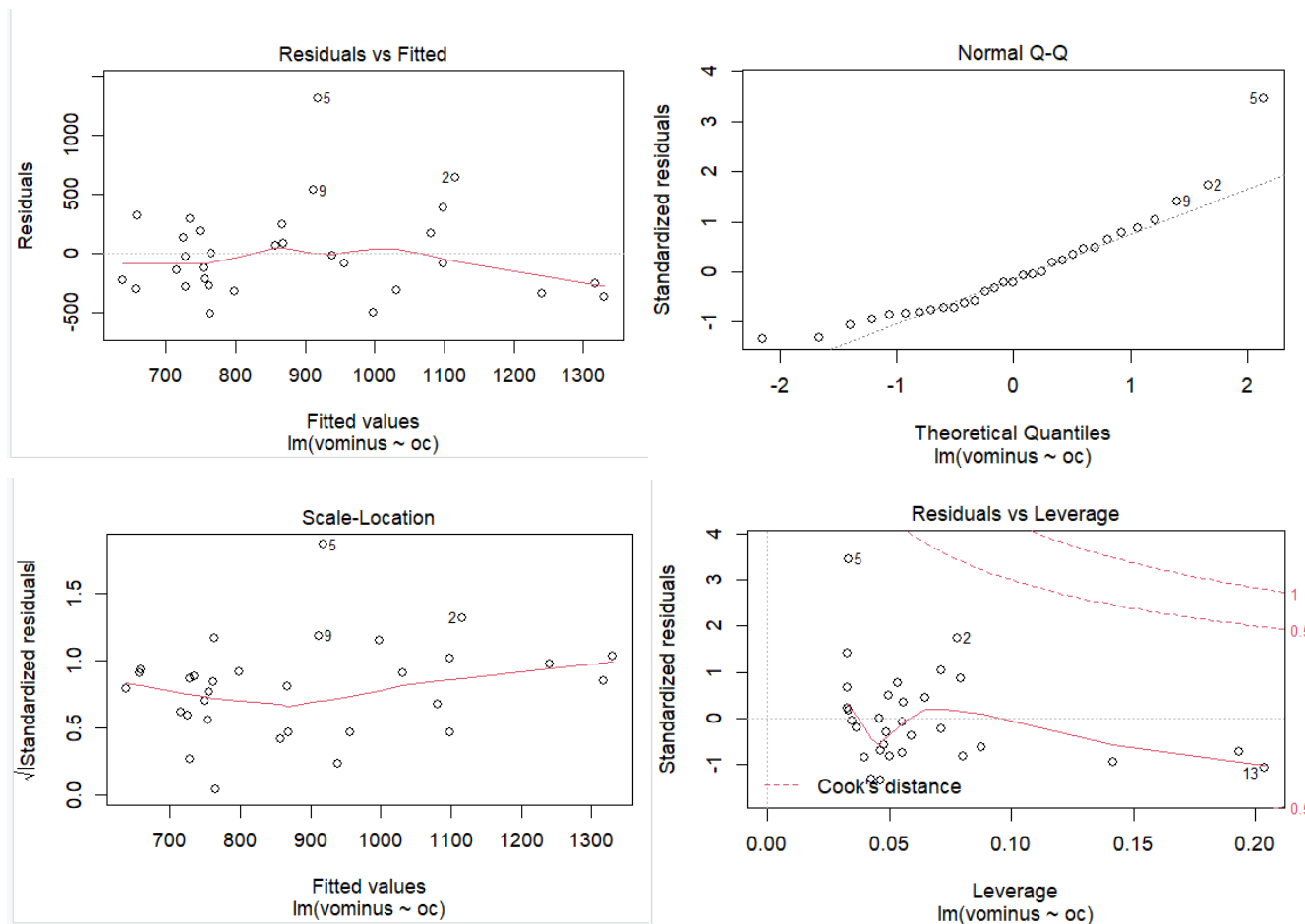$$\widehat{VO-} = 309.051 - 1.868\,OC + 48.501\,TRAP$$

Residual standard error: 324.4 on 28 degrees of freedom
Multiple R-squared: 0.463,   Adjusted R-squared: 0.4247
F-statistic: 12.07 on 2 and 28 DF, p-value: 0.0001658

This view is consistent with the pattern of relationships described in the previous exercise, as TRAP has a stronger correlation with VO- than OC does. The coefficient of TRAP is also much greater than the coefficient of OC, meaning it is more significant when both are present in the model. Trap also has a lower p value, which means it is a better predictor.

```
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 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 324.4 on 28 degrees of freedom
Multiple R-squared:  0.463,     Adjusted R-squared:  0.4247
F-statistic: 12.07 on 2 and 28 DF,  p-value: 0.0001658

Analysis of Variance Table

Response: vominus
          Df  Sum Sq Mean Sq F value   Pr(>F)
oc         1 1134474 1134474  10.784 0.002752 **
trap       1 1405460 1405460  13.359 0.001051 **
Residuals 28 2945733  105205
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$VO- = \beta_0 + \beta_1 OC + \beta_2 TRAP + \beta_3\, VO+ + \varepsilon_i\ i = 1,\dots,n$$
Assumption: ε are independent $N(0,\sigma)$ variables.

Estimated equation: VO- = 267.26110 + 0.72420 VO+ + 9.48453 trap -6.51323 OC + ε_i

Due to the larger p-value (0.29), trap is not significant.

Residual standard error: 179.2 on 27 degrees of freedom
Multiple R-squared: 0.842,   Adjusted R-squared: 0.8245
F-statistic: 47.97 on 3 and 27 DF, p-value: 5.974e-11

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 267.26110   74.71782   3.577  0.00134 **
voplus        0.72420    0.08999   8.048  1.2e-08 ***
oc           -6.51323    2.50744  -2.598  0.01502 *
trap          9.48453    8.78782   1.079  0.29001
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 179.2 on 27 degrees of freedom
Multiple R-squared:  0.842,     Adjusted R-squared:  0.8245
F-statistic: 47.97 on 3 and 27 DF,  p-value: 5.974e-11

Analysis of Variance Table

Response: vominus
          Df  Sum Sq Mean Sq  F value    Pr(>F)
voplus     1 4401727 4401727 137.1303 4.332e-12 ***
oc         1  179880  179880   5.6039   0.02534 *
trap       1   37390   37390   1.1648   0.29001
Residuals 27  866669   32099
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model	Variable	Variable	Variable t-statistic	Variable p-value	R^2	S

	Coefficients	Standard Error				
OC	OC: 9.917	OC: 3.606	OC: 2.75	OC: p < 0.05	0.2068	387.4
OC, TRAP	OC: -1.868 TRAP:48. 501	OC: 4.418 TRAP: 13.27	OC: -0.42 TRAP: 3.655	OC: 0.676 TRAP: 0.00106	0.463	324.4
OC, TRAP, VO+	OC: -6.51323 TRAP: 9.484 VO+: 0.72420	OC: 2.5 TRAP: 8.788 VO+: 0.09	OC: -2.6 TRAP: 1.079 VO+: 8.048	OC: 0.015 TRAP: 0.29 VO+: 1.2e-8	0.842	179.2

- OC:
  - Variation: Multiple R^2 = 0.2068 = 20.68%
  - Residual standard error=σ= 387.4
- OC, TRAP:
  - Variation: Multiple R^2 = 0.463 = 46.3%
  - Residual standard error = σ= 324.4
- OC,TRAP, VO-:
  - Variation: Multiple R^2 = 0.842= 84.2%
  - σ = 179.2

- Summary: Percent of variation in VO- explained by OC is the lowest, but the standard error is the highest. Percent of variation in VO- explained by OC, Trap, and VO+ is the highest, but the standard error is the lowest.

Model without trap:

Estimated equation: VO- = 298.01211 +0.778VO+ -5.25 OC + ε_i

Residual standard error: 179.7 on 28 degrees of freedom
Multiple R-squared: 0.8352,   Adjusted R-squared: 0.8234
F-statistic: 70.95 on 2 and 28 DF, p-value: 1.09e-11

Variation for estimation without using trap is slightly lower than estimation with trap (0.835< 0.842), and standard error of estimation without using trap is slightly larger than the one in part (b) (179.7>179.2). The values are very similar.

```
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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 179.7 on 28 degrees of freedom
Multiple R-squared:  0.8352,    Adjusted R-squared:  0.8234
F-statistic: 70.95 on 2 and 28 DF,  p-value: 1.09e-11

Analysis of Variance Table

Response: vominus
          Df  Sum Sq Mean Sq F value    Pr(>F)
oc         1 1134474 1134474  35.136 2.224e-06 ***
voplus     1 3447133 3447133 106.763 4.651e-11 ***
Residuals 28  904060   32288
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

11.41 Predicting bone resorption using transformed variables.

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.2110     0.4161   12.52  3.2e-13 ***
loc           0.4406     0.1227    3.59   0.0012 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4089 on 29 degrees of freedom
Multiple R-squared:  0.3077,    Adjusted R-squared:  0.2838
F-statistic: 12.89 on 1 and 29 DF,  p-value: 0.001202

Analysis of Variance Table

Response: lvominus
          Df Sum Sq Mean Sq F value    Pr(>F)
loc        1 2.1556 2.15560  12.889 0.001202 **
Residuals 29 4.8499 0.16724
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
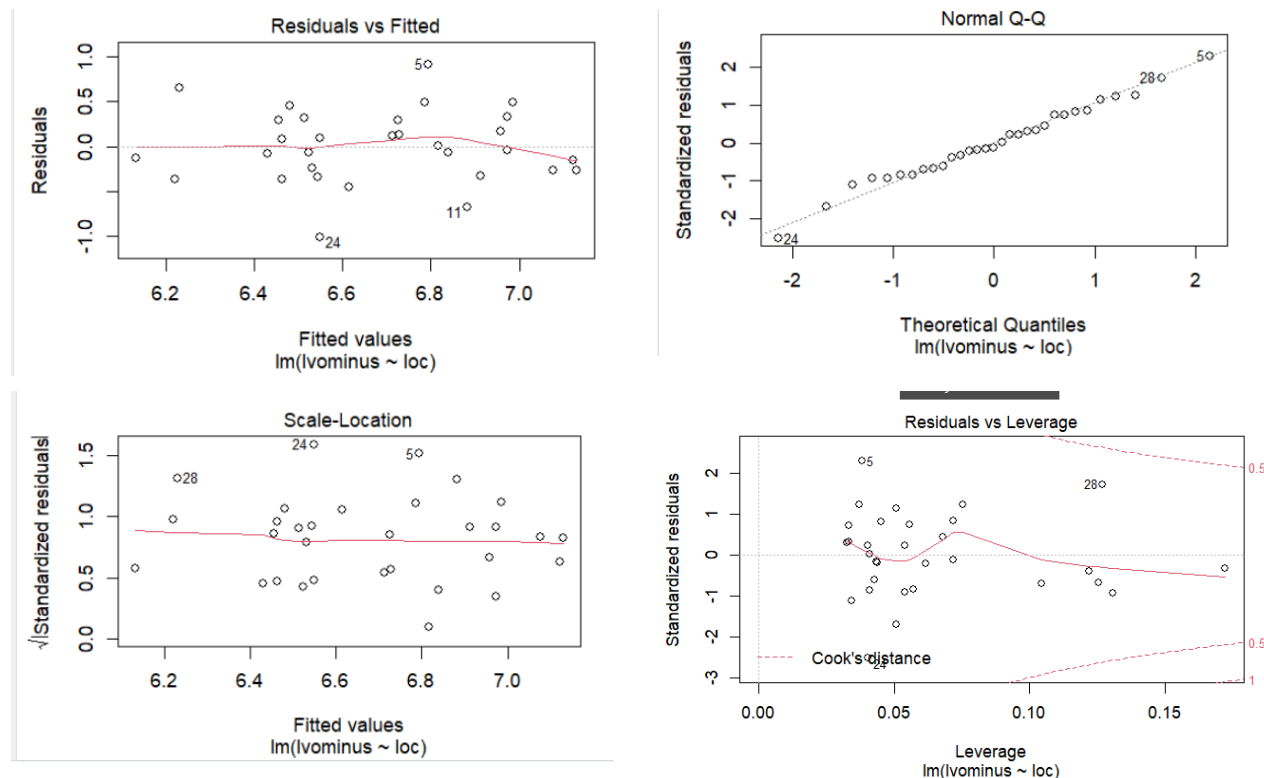
$$\widehat{LVO-} = 5.2110 + 0.4406\,LOC$$

Residual standard error: 0.4089 on 29 degrees of freedom

Multiple R-squared: 0.3077,   Adjusted R-squared: 0.2838

F-statistic: 12.89 on 1 and 29 DF, p-value: 0.001202

T value = 3.59

For LTRAP and LOC

$$\widehat{LVO-} = 5.03705 + 0.05688\,LOC + 0.58956L\,TRAP$$

Residual standard error: 0.3732 on 28 degrees of freedom

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

F-statistic: 11.14 on 2 and 28 DF, p-value: 0.0002754

```
      Min      1Q   Median      3Q      Max
 -1.04753 -0.23066  0.02055  0.24900  0.73127

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  5.03705    0.38562  13.062 1.96e-13 ***
 loc          0.05688    0.18480   0.308   0.7605
 ltrap        0.58956    0.22588   2.610   0.0144 *
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 0.3732 on 28 degrees of freedom
 Multiple R-squared:  0.4432,    Adjusted R-squared:  0.4034
 F-statistic: 11.14 on 2 and 28 DF,  p-value: 0.0002754

 Analysis of Variance Table

 Response: lvominus
           Df Sum Sq Mean Sq F value   Pr(>F)
 loc        1 2.1556 2.15560 15.4728 0.000502 ***
 ltrap      1 0.9491 0.94910  6.8126 0.014373 *
 Residuals 28 3.9008 0.13932
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For All:

$$\widehat{LVO-} = 1.5729 - 0.2932LOC + 0.2447L\,TRAP + 0.8134\,LVO +$$

Residual standard error: 0.2558 on 27 degrees of freedom

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

F-statistic: 26.68 on 3 and 27 DF, p-value: 3.132e-08

```
 Call:
 lm(formula = lvominus ~ loc + ltrap + lvoplus, data = project)

 Residuals:
      Min      1Q   Median      3Q      Max
 -0.79210 -0.09672  0.01608  0.15194  0.38681

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)   1.5729     0.6618   2.377   0.0248 *
 loc          -0.2932     0.1407  -2.084   0.0468 *
 ltrap         0.2447     0.1662   1.472   0.1525
 lvoplus       0.8134     0.1425   5.710 4.56e-06 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 0.2558 on 27 degrees of freedom
 Multiple R-squared:  0.7478,    Adjusted R-squared:  0.7197
 F-statistic: 26.68 on 3 and 27 DF,  p-value: 3.132e-08

 Analysis of Variance Table

 Response: lvominus
           Df Sum Sq Mean Sq F value   Pr(>F)
 loc        1 2.1556 2.15560  32.936 4.22e-06 ***
 ltrap      1 0.9491 0.94910  14.502 0.0007337 ***
 lvoplus    1 2.1337 2.13372  32.602 4.56e-06 ***
 Residuals 27 1.7671 0.06545
```

Without LTRAP:

$$\widehat{LVO-} = 1.3109 - 0.1878\,LOC + 0.8896\,LVO +$$

Residual standard error: 0.2611 on 28 degrees of freedom

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

F-statistic: 37.38 on 2 and 28 DF, p-value: 1.245e-08

```
 Call:
 lm(formula = lvominus ~ loc + lvoplus, data = project)

 Residuals:
      Min      1Q   Median      3Q      Max
 -0.75504 -0.04458  0.03341  0.16852  0.36422

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)   1.3109     0.6506   2.015   0.0536 .
 loc          -0.1878     0.1237  -1.519   0.1400
 lvoplus       0.8896     0.1355   6.568 4.03e-07 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 0.2611 on 28 degrees of freedom
 Multiple R-squared:  0.7275,    Adjusted R-squared:  0.708
 F-statistic: 37.38 on 2 and 28 DF,  p-value: 1.245e-08

 Analysis of Variance Table

 Response: lvominus
           Df Sum Sq Mean Sq F value   Pr(>F)
 loc        1 2.1556 2.15560  31.617 5.072e-06 ***
 lvoplus    1 2.9409 2.94094  43.136 4.028e-07 ***
 Residuals 28 1.9090 0.06818
```