Report on Relationship between Bone Formation and Resorption

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

Section 1: Report

Executive Summary

This reports summarizes our findings on producing a model which can best predict bone resorption and formation. The result of this model shows that the best single indicator for bone resorption and formation is the two variables $VO-$ against $VO+$ that measure the two respectively. While the best multiple regression indicator was the prediction for $VO+$ against $OC, trap, VO-$. Thus, $VO-$ against $VO+$ are strong single predictors, while $VO+$ and $OC, trap, VO-$ are strong multiple regression predictors for measuring the bone renewal process in healthy woman between the ages of 11 to 32 years.

Data Set

The dataset contains four primary variables related to bone renewal, this includes two measurements previously mentioned $VO-$ and $VO+$, and two biomarkers Osteocalcin (OC) and tartrate-resistant acid phosphatase ($trap$). The sample size of $n = 31$ comes from measurements in a study of 31 healthy woman aged 11 to 32 years. We make two versions of the study, where one is with the original dataset, and another where the data is log transformed. This leads to a normalization of data from the significant right skew seen in the data due to large max values.

Software

Our analysis was performed entirely in R (4.1.0), via RMarkdown, in conjunction with LaTeX to produce the research paper.

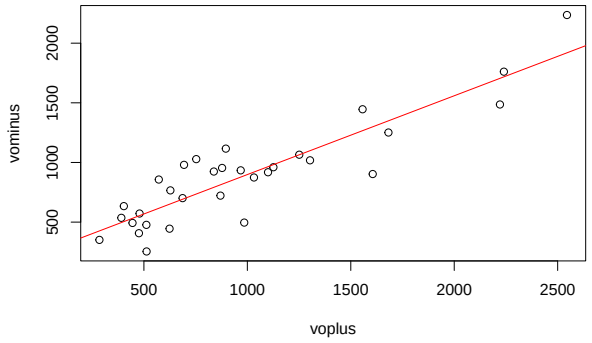
Methodology

The approach used for this study was linear regression analysis, where for the first portion we predicted $VO+$ against the biomarkers OC, then both $OC, trap$ to see if there was a relationship between measurements and biomarkers. Then we used $VO+$ against all three variables $OC, trap, VO-$ to see which out of the three varaibles would likely be a strong predictior of measurements. This analysis was repeated again for log transformed data, and the two parts were repeated with the $VO-$ measurement instead. The goal of the analysis was to find strong linear relationships between bone resorption and formation.

Results

For Single Linear Regression:

$VO-$ vs $VO+$




```
##                  2.5 %      97.5 %
## (Intercept) 96.1866939 379.7809145
## voplus       0.5360875   0.7850841
```

```
##  R-squared Adjusted R-squared Estimate std
##     0.8024             0.7956     193.3319
```

$vominus = 237.984 + 0.661(voplus)$

Above is the final result for the best single linear regression in our study, with the 95% confidence intervals for the regression line. The data shows that there is a strong linear relationship ($R^2 = .8024$) between $VO+$ and $VO-$, and that further data may lead to similar and more vigorous results.

For Multiple Linear Regression:

$VO+$ vs $OC, trap, VO-$

(Data is 4-dimensional)

```
##                    2.5 %     97.5 %
## (Intercept) -436.8076675 -50.167750
## oc             2.4083935  14.061364
## trap         -14.5963816  27.810667
## vominus        0.7261057   1.223037
```

```
##  R-squared Adjusted R-squared Estimate std
##     0.8844             0.8715     207.8362
```

$voplus = -243.488 + 8.235(oc) + 6.607(trap) + 0.975(vominus)$

Above is the final result for the best multiple regression in our study, with the 95% confidence intervals for the regression line. The data shows a more convincing result, ($R^2 = .8844$), further proves that with more data this model can be a strong predictor for bone formation ($VO+$).

Other Results:

As seen in Section 2 (details of study), other than the above results, R^2 tend to lie between $.4 < R^2 < .6$ range, noteably models with less variables, and log transformed data result in vastly lower R^2. The worst regression model being log transformed $VO-$ vs OC with a $R^2 = 0.3076$

Conclusion

Although there is a very limited amount of data, the results indicate strong linear relationships between a large portion of the regression analysis done, notability the two identified in the Results section. To reduce variability, further studies should be done to collect more data.

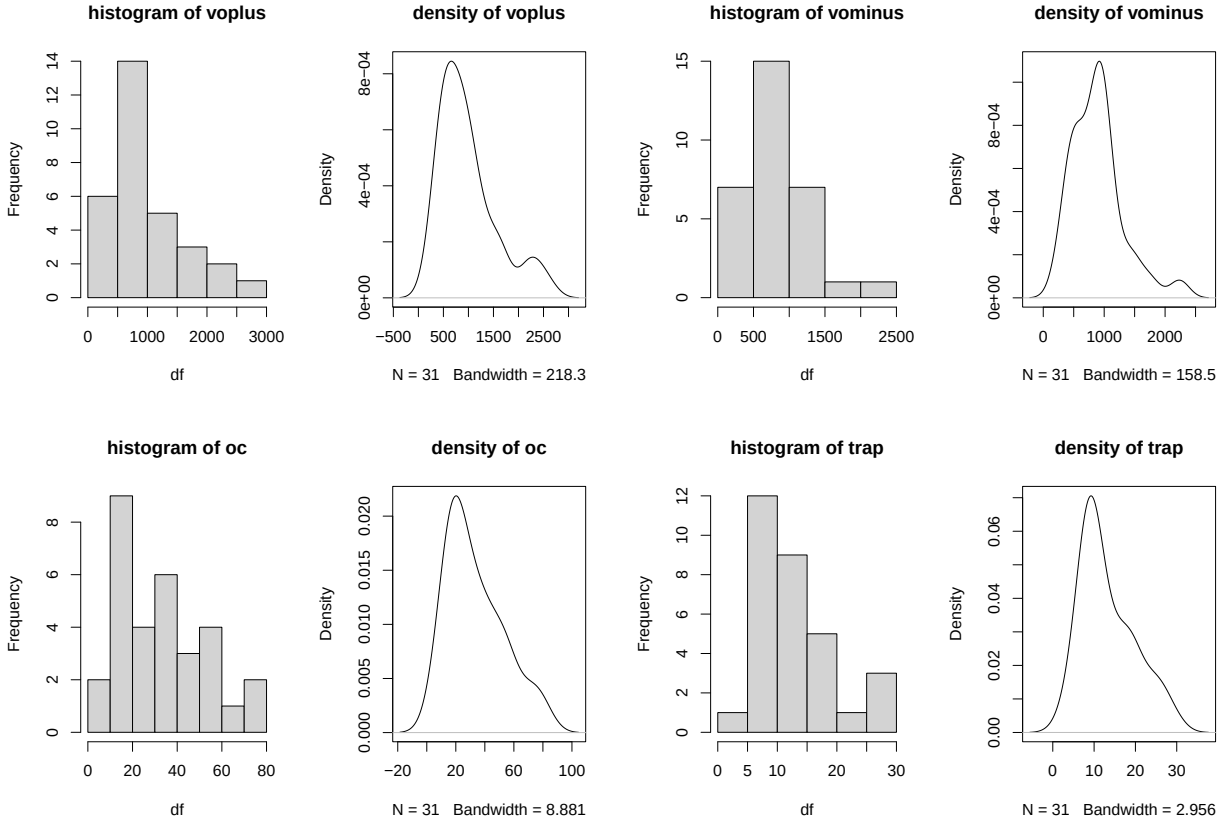
Since the study that led to the data tells us that healthy bones specifically go through bone renewal, we can overlook the health status of a subject when it comes to variability in data. But to expand this model towards a larger audience, data should be sampled from a larger age range ($\pm$ 5 years), and samples should be taken from men as well. This should lead to a model which has more sturdiness, and usability when it comes to studying bone renewal.

Section 2: Details of Study

11.36 Bone formation and resorption

a)

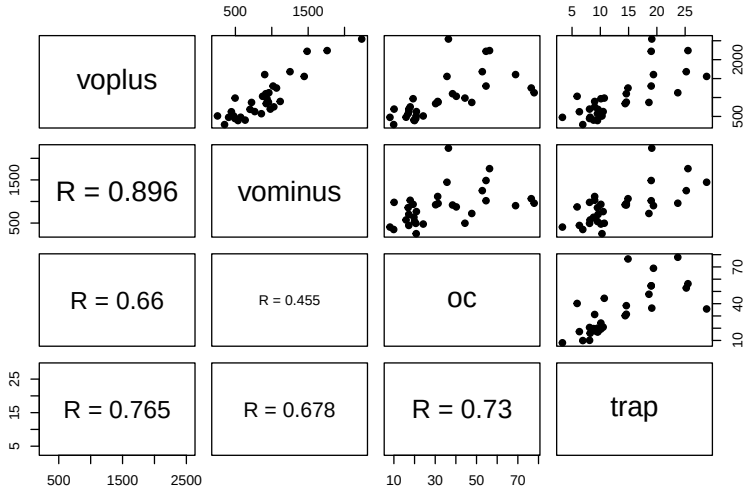
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std	IQR
voplus	285.0	542.5	870.0	985.806	1188.50	2545.0	579.858	646.00
vominus	254.0	554.0	903.0	889.194	1023.00	2236.0	427.616	469.00
oc	8.1	18.6	30.2	33.416	46.05	77.9	19.610	27.45
trap	3.3	8.9	10.3	13.248	18.80	28.8	6.528	9.90



All four variables $VO+, VO-, OC$ and *trap* have a similar right skew in the histogram. Their density graphs shows the distributions of data are similar as well.

Although it appears in all four plots, they have an outlier based on their respective maxes, as seen in the summary.

b)



The pairs show that there is a strong positive correlation between $VO+$ and $VO-$, (r = .896), and a high positive correlation between $VO+, OC$ and *trap*, (r = .765, .73)

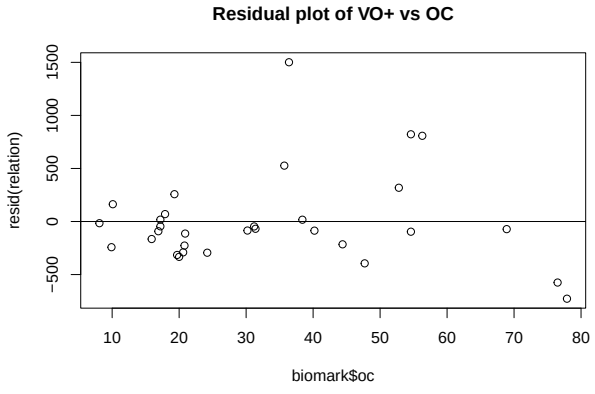
There are moderately positive correlations between $VO+$ and OC, and in $VO+$ and *trap* ($0.65 < r < .7$)

11.37 Predicting bone formation

Let's use regression methods to predict VO+, the measure of bone formation.

a) Simple Linear Regression: VO+ vs OC

```
## [1] "Short Regression Summary:"
## lm(formula = voplus ~ oc, data = biomark)
##             Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 334.03439 159.240942 2.097667 0.04476326
## oc           19.50471   4.127054 4.726062 0.00005429
## [1] "F-statistic: 22.3356625977023"
```



b)

```
## [1] "Short Regression Summary:"
## lm(formula = voplus ~ oc + trap, data = biomark)
##             Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 57.704192 156.538826 0.3686254 0.715179551
## oc           6.414658   5.124554 1.2517495 0.221017683
## trap        53.874424  15.393305 3.4998607 0.001577044
```

```
## [1] "F-statistic: 21.6243163032774"
```

Fitted regression: $voplus = 57.704 + 6.415(oc) + 53.874(trap)$

OC Hypothesis Test:

$H_0 : \beta_1 = 0$ (no significant diff. from zero)

$H_a : \beta_1 \neq 0$ (significant diff. from zero)

As we can see the P-value for OC is larger than significance level 0.05. Thus, we FAIL TO REJECT the null hypothesis. Meaning, that there is no significant difference.

trap Hypothesis Test:

$H_0 : \beta_2 = 0$ (no significant diff. from zero)

$H_a : \beta_2 \neq 0$ (significant diff. from zero)

As we can see the P-value for TRAP is much less than significance level 0.05. Thus, we REJECT the null hypothesis. Meaning, that there is a significant difference.

The above results tell us that TRAP is measured with more precision than OC. This is consistent with the correlation found in exercise 11.36.

11.38 More on predicting bone formation.

Now consider a regression model for predicting VO+ using OC, TRAP, and VO-.

a)

$voplus = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 z_i + \epsilon_i,\ i = 1, \cdots n$

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

s.t $(x, y, z) = (oc, trap, vominus)$

b)

```
##
## Call:
## lm(formula = voplus ~ oc + trap + vominus, data = df)
##
## 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 *
## 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 ***
## ---
## 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
```

$voplus = -243.488 + 8.235(oc) + 6.607(trap) + 0.975(vominus)$

c)

coefficients and the P-values differ for the three analyses.

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -243.4877089 94.2182880 -2.5842935 1.548672e-02
## oc             8.2348785  2.8396522  2.8999603 7.331775e-03
## trap           6.6071427 10.3339550  0.6393624 5.279750e-01
## vominus        0.9745712  0.1210945  8.0480191 1.198851e-08
```

```
##             Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 334.03439 159.240942 2.097667 0.04476326
## oc           19.50471   4.127054 4.726062 0.00005429
```

```
##             Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 57.704192 156.538826 0.3686254 0.715179551
## oc           6.414658   5.124554 1.2517495 0.221017683
## trap        53.874424  15.393305 3.4998607 0.001577044
```

OC and *trap* both have higher p-values in the multiple regression model, compared to the previous models. There is much higher p-value when testing $VO+$ and *OC* on its own.

The coefficients are much more normalized in the multiple regression model.

d)

```
##  R-squared Adjusted R-squared Estimate std
##     0.8844             0.8715     207.8362

##  R-squared Adjusted R-squared Estimate std
##     0.4351             0.4156     443.2745

##  R-squared Adjusted R-squared Estimate std
##      0.607             0.5789     376.2652
```

The higher R^2 in the multiple regression model indicates it is much better of a model compared to the other two.

e)

The results from part b, show that there is a very low p-value for vominus, thus there is strong evidence that a model between voplus and vominus will result in an effective model.

```
##
## Call:
## lm(formula = voplus ~ vominus, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -402.11 -176.16  -47.07  124.01  603.42
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -94.2855   110.1056  -0.856    0.399
## vominus       1.2147     0.1119  10.852    1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 262.2 on 29 degrees of freedom
## Multiple R-squared:  0.8024, Adjusted R-squared:  0.7956
## F-statistic: 117.8 on 1 and 29 DF,  p-value: 1e-11

##              Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept) -94.285549 110.1055916 -0.8563194 3.988419e-01
## vominus       1.214687   0.1119326 10.8519548 1.000370e-11

##  R-squared Adjusted R-squared Estimate std
##     0.8024             0.7956     262.1629
```

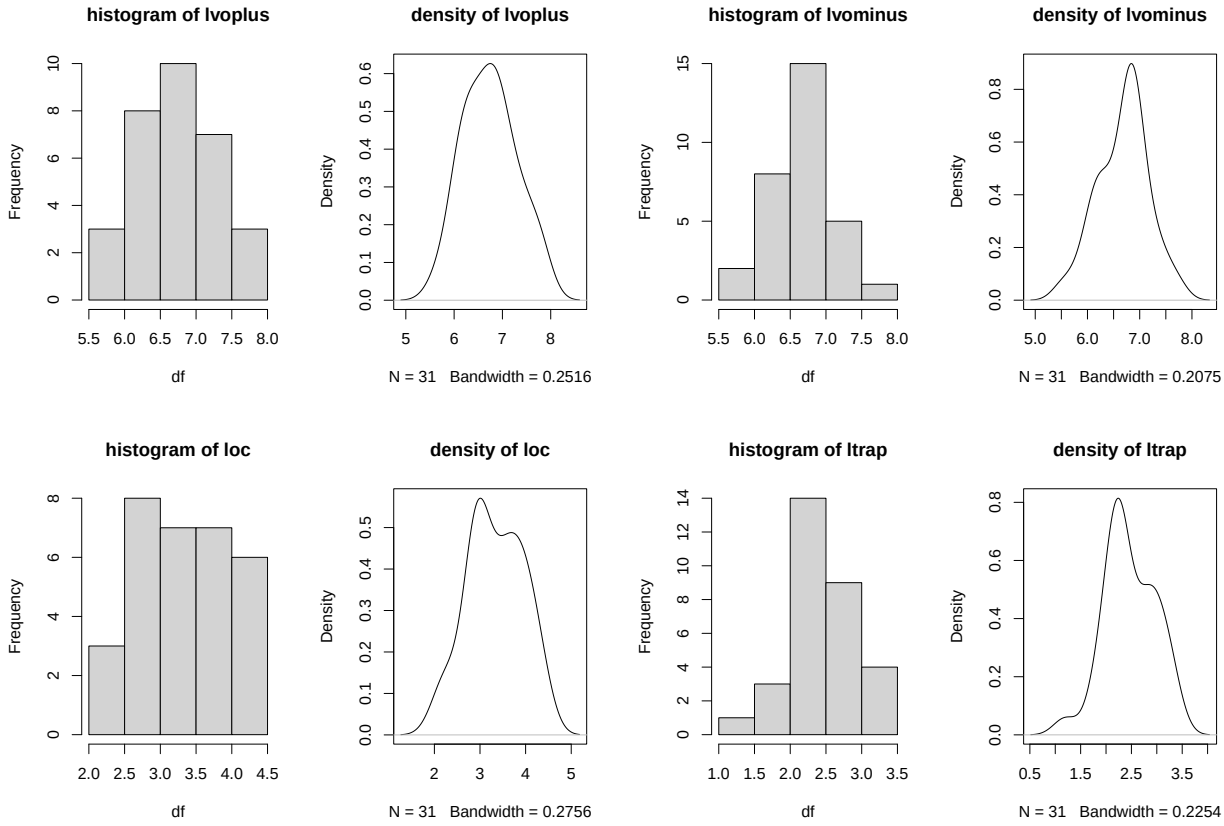
In comparison with the multiple regression model, a model based on voplus and vominus has a much lower p-value for vominus, although the p-value for intercept is a power of ten higher. Further proving there is a strong relation between voplus and vominus.

11.39 Predicting bone formation using transformed variables.

Repeat above parts, but transform data via $log(Biomark)$

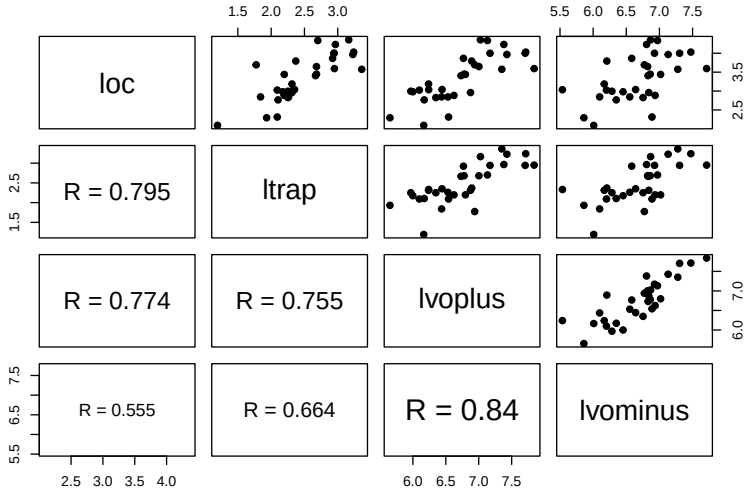
1a)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std	IQR
lvoplus	5.652	6.295	6.768	6.742	7.079	7.842	0.555	0.784
lvominus	5.537	6.316	6.806	6.682	6.931	7.712	0.483	0.614
loc	2.092	2.922	3.408	3.338	3.829	4.355	0.609	0.907
ltrap	1.194	2.186	2.332	2.467	2.934	3.360	0.498	0.748



$VO+$ is relatively normal, $VO-$ is slightly right skewed, and both OC and *trap* have a left skew, where trap is more skewed. After applying logs to the biomarkers, the data appears to be much more normally distributed compared to the original data.

1b)

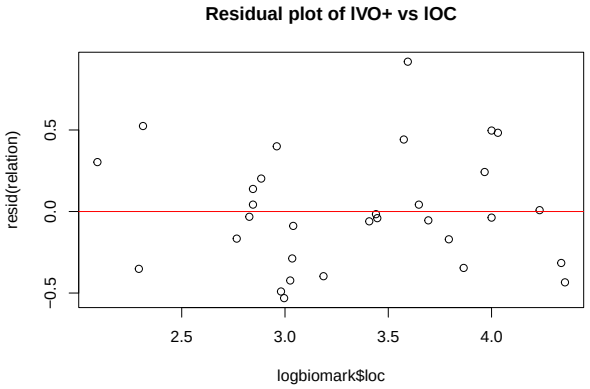


The pairs show that there is a strong positive correlation between $VO+$ and $VO-$, (r = .84), and a high positive correlation between $VO+, trap$ and OC (r = .774, .795) and $VO+$ and trap (r = .775)

There are moderately positive correlations between $VO-+$ and *trap* ($0.65 < \text{r} < .7$)

2a)

```
## [1] "Short Regression Summary:"
## lm(formula = logbiomark$lvoplus ~ logbiomark$loc)
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    4.3846345  0.3642858 12.036249 8.420228e-13
## logbiomark$loc 0.7061896  0.1074218  6.573988 3.341897e-07
## [1] "F-statistic: 43.2173149698603"
```



2b)

```
## [1] "Short Regression Summary:"
## lm(formula = lvoplus ~ loc + ltrap, data = logbiomark)
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 4.2592427  0.3506399 12.147056 1.117053e-12
## loc         0.4301326  0.1680072  2.560204 1.614440e-02
## ltrap       0.4242738  0.2053691  2.065908 4.820199e-02
## [1] "F-statistic: 26.1777014322833"
```

Fitted regression: $lvoplus = 4.2592 + 0.4301(loc) + 0.4243(ltrap)$

LOC Hypothesis Test:

$H_0 : \beta_1 = 0$ (no significant diff. from zero)

$H_a : \beta_1 \neq 0$ (significant diff. from zero)

As we can see the P-value for LOC is less than significance level 0.05. Thus, we REJECT the null hypothesis. Meaning, that there is a significant difference.

ltrap Hypothesis Test:

$H_0 : \beta_2 = 0$ (no significant diff. from zero)

$H_a : \beta_2 \neq 0$ (significant diff. from zero)

As we can see the P-value for LTRAP is less than significance level 0.05. Thus, we REJECT the null hypothesis. Meaning, that there is a significant difference.

The above results tell us that both LOC and LTRAP have a significant difference. LOC is measured with more precision than LTRAP.

This is not consistent with the correlation found in exercise 11.36.

3a)

Given logarithm applied to dataset:

$lvoplus = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 z_i + \epsilon_i,\ i = 1, \cdots n$

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

s.t $(x, y, z) = (loc, ltrap, lvominus)$

3b)

```
##
## Call:
## lm(formula = lvoplus ~ loc + ltrap + lvominus, data = logbiomark)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -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 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2326 on 27 degrees of freedom
## Multiple R-squared:  0.8421, Adjusted R-squared:  0.8246
## F-statistic: 48.02 on 3 and 27 DF,  p-value: 5.906e-11
```

$lvoplus = 0.872 + 0.392(loc) + 0.028(ltrap) + 0.673(lvominus)$

3c)

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.87152677  0.6401485 1.3614447 1.846273e-01
## loc         0.39196914  0.1153479 3.3981483 2.120342e-03
## ltrap       0.02768179  0.1569668 0.1763544 8.613316e-01
## lvominus    0.67254275  0.1177885 5.7097480 4.560821e-06

##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    4.3846345  0.3642858 12.036249 8.420228e-13
## logbiomark$loc 0.7061896  0.1074218  6.573988 3.341897e-07

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 4.2592427  0.3506399 12.147056 1.117053e-12
## loc         0.4301326  0.1680072  2.560204 1.614440e-02
## ltrap       0.4242738  0.2053691  2.065908 4.820199e-02
```

When applying log transformation, the first model has a much lower p-value, compared to the multiple regression and second model. This may indicate it is a possible choice for a valid model.

The coefficients in the multiple regression model are more bounded between 0-1.

3d)

```
##  R-squared Adjusted R-squared Estimate std
##     0.8421             0.8246       0.2326

##  R-squared Adjusted R-squared Estimate std
##     0.5984             0.5846        0.358

##  R-squared Adjusted R-squared Estimate std
##     0.6515             0.6267       0.3394
```

The R^2 shows us that the multiple regression model, seems to be the better fit.

3e)

The results from part b, shows when applying logarithm to the data, that there still is a very low p-value for vominus, thus there is strong evidence that a model between voplus and vominus will result in an effective model.

```
##
## Call:
## lm(formula = lvoplus ~ lvominus, data = logbiomark)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52124 -0.25425  0.03444  0.15993  0.60924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2926     0.7765   0.377    0.709
## lvominus      0.9652     0.1159   8.326 3.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3068 on 29 degrees of freedom
## Multiple R-squared:  0.7051, Adjusted R-squared:  0.6949
## F-statistic: 69.32 on 1 and 29 DF,  p-value: 3.536e-09

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.2925777  0.7765489 0.3767667 7.090899e-01
## lvominus    0.9652298  0.1159293 8.3260219 3.535857e-09
```

```
##  R-squared Adjusted R-squared Estimate std
##     0.7051            0.6949       0.3068
```

In comparison with the multiple regression model, a model based on voplus and vominus has a much lower p-value for vominus, although the p-value for intercept is a power of ten higher. Further proving there is a strong relation between voplus and vominus.

11.40 Predicting bone resorption.

To predict bone resperation, reverse the roles of VO+ and VO-

1a)

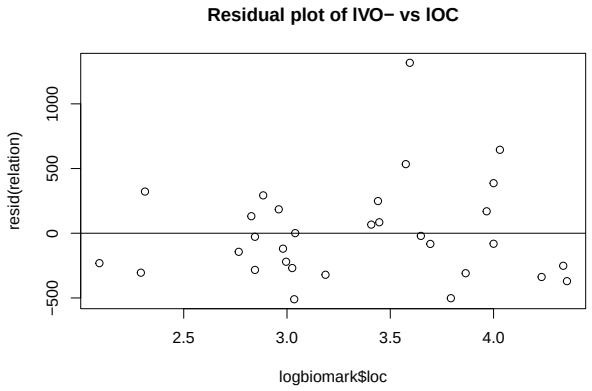
Refer to 11.39 for log numerical analysis

1b)

Refer to 11.39 for log pairs

2a)

```
## [1] "Short Regression Summary:"
## lm(formula = biomark$vominus ~ biomark$oc)
##                Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 557.817691 139.151268 4.008714 0.000390706
## biomark$oc    9.916644   3.606389 2.749743 0.010161429
## [1] "F-statistic: 7.56108686083209"
```



2b)

```
## [1] "Short Regression Summary:"
## lm(formula = biomark$vominus ~ biomark$oc + biomark$trap)
##                Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)  309.050693 134.941549  2.2902560 0.029744840
## biomark$oc    -1.867714   4.417532 -0.4227959 0.675673633
## biomark$trap  48.500594  13.269529  3.6550351 0.001050919
## [1] "F-statistic: 12.0713842009958"
```

Fitted regression: $vominus = 309.051 + -1.868(oc) + 48.501(trap)$

OC Hypothesis Test:

$H_0 : \beta_1 = 0$ (no significant diff. from zero)

$H_a : \beta_1 \neq 0$ (significant diff. from zero)

As we can see the P-value for OC is much larger than significance level 0.05. Thus, we FAIL TO REJECT the null hypothesis. Meaning, that there is not a significant difference.

trap Hypothesis Test:

$H_0 : \beta_2 = 0$ (no significant diff. from zero)

$H_a : \beta_2 \neq 0$ (significant diff. from zero)

As we can see the P-value for TRAP is much less than significance level 0.05. Thus, we REJECT the null hypothesis. Meaning, that there is a significant difference.

The above results tell us that TRAP is measured with more precision than OC.

This is consistent with the correlation found in exercise 11.36.

3a)

$vominus = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 z_i + \epsilon_i,\ i = 1, \cdots n$

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

s.t $(x, y, z) = (oc, trap, voplus)$

3b)

```
##
## Call:
## lm(formula = vominus ~ oc + trap + voplus, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -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 ***
## ---
## 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
```

$vominus = 267.261 + -6.513(oc) + 9.485(trap) + 0.724(voplus)$

3c)

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 267.2610974 74.71781543  3.576939 1.339554e-03
## oc           -6.5132341  2.50744021 -2.597563 1.501820e-02
## trap          9.4845308  8.78782444  1.079281 2.900101e-01
## voplus        0.7242038  0.08998535  8.048019 1.198851e-08

##               Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)  309.050693 134.941549  2.2902560 0.029744840
## biomark$oc    -1.867714   4.417532 -0.4227959 0.675673633
## biomark$trap  48.500594  13.269529  3.6550351 0.001050919

##             Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 4.2592427  0.3506399 12.147056 1.117053e-12
## loc         0.4301326  0.1680072  2.560204 1.614440e-02
## ltrap       0.4242738  0.2053691  2.065908 4.820199e-02
```

Swapping the prediction values, gives a familiar result to the models made in 11.37-11.38, the lower p-values in the multiple regression model overall having a stronger model.

The coefficients are similar results to that in 11.37-11.38

3d)

```
##  R-squared Adjusted R-squared Estimate std
##      0.842             0.8245     179.1616

##  R-squared Adjusted R-squared Estimate std
##      0.463             0.4247     324.3528

##  R-squared Adjusted R-squared Estimate std
##     0.6515             0.6267       0.3394
```

The R^2 shows the multiple regression model is likely a good fit

3e)

The results from part b, show that there is a very low p-value for voplus, thus there is strong evidence that a model between voplus and vominus will result in an effective model.

```
##
## Call:
## lm(formula = vominus ~ voplus, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.88  -94.39    0.62  131.12  316.83
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 237.98380   69.33065   3.433  0.00182 **
## voplus        0.66059    0.06087  10.852    1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193.3 on 29 degrees of freedom
## Multiple R-squared:  0.8024, Adjusted R-squared:  0.7956
## F-statistic: 117.8 on 1 and 29 DF,  p-value: 1e-11

##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 237.9838042 69.33065480  3.432591 1.819033e-03
## voplus        0.6605858  0.06087252 10.851955 1.000370e-11

##  R-squared Adjusted R-squared Estimate std
##     0.8024             0.7956     193.3319
```

In comparison with the multiple regression model, a model based on voplus and vominus has a much lower p-value for vominus, although the p-value for intercept is a power of ten higher. Further proving there is a strong relation between voplus and vominus.

11.41 Predicting bone resorption using transformed variables.

Repeat above exercise, but transform data via $log(Biomark)$

1a)

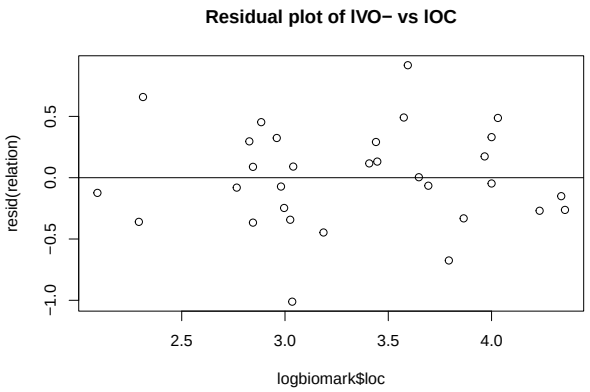
Refer to 11.39 for log numerical analysis

1b)

Refer to 11.39 for log pairs

2a)

```
## [1] "Short Regression Summary:"
## lm(formula = logbiomark$lvominus ~ logbiomark$loc)
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    5.2114553  0.4161271 12.523712 3.187292e-13
## logbiomark$loc 0.4404316  0.1227089  3.589238 1.204697e-03
## [1] "F-statistic: 12.882632846463"
```



2b)

```
## [1] "Short Regression Summary:"
## lm(formula = logbiomark$lvominus ~ logbiomark$loc + logbiomark$ltrap)
##                     Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)      5.03717552  0.3855945 13.0634019 1.958962e-13
## logbiomark$loc   0.05674508  0.1847554  0.3071362 7.610124e-01
## logbiomark$ltrap 0.58969044  0.2258420  2.6110756 1.433985e-02
## [1] "F-statistic: 11.1423725427422"
```

Fitted regression: $vominus = 5.03718 + 0.05675(loc) + 0.58969(ltrap)$

OC Hypothesis Test:

$H_0 : \beta_1 = 0$ (no significant diff. from zero)

$H_a : \beta_1 \neq 0$ (significant diff. from zero)

As we can see the P-value for LOC is larger than significance level 0.05. Thus, we FAIL TO REJECT the null hypothesis. Meaning, that there is not a significant difference.

trap Hypothesis Test:

$H_0 : \beta_2 = 0$ (no significant diff. from zero)

$H_a : \beta_2 \neq 0$ (significant diff. from zero)

As we can see the P-value for LTRAP is less than significance level 0.05. Thus, we REJECT the null hypothesis. Meaning, that there is a significant difference.

The above results tell us that LTRAP is measured with more precision than LOC.

This is consistent with the correlation found in exercise 11.36.

3a)

Given logarithm applied to dataset:

$vominus = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 z_i + \epsilon_i,\ i = 1, \cdots n$

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

s.t $(x, y, z) = (oc, trap, voplus)$

3b)

```
##
## Call:
## lm(formula = vominus ~ oc + trap + voplus, data = log(df))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79189 -0.09669  0.01564  0.15182  0.38662
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5728     0.6620   2.376   0.0249 *
## oc           -0.2932     0.1407  -2.084   0.0468 *
## trap          0.2448     0.1662   1.473   0.1523
## voplus        0.8134     0.1425   5.709 4.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2559 on 27 degrees of freedom
## Multiple R-squared:  0.7477, Adjusted R-squared:  0.7196
## F-statistic: 26.67 on 3 and 27 DF,  p-value: 3.147e-08
```

$vominus = 1.573 + -0.293(oc) + 0.245(trap) + 0.813(voplus)$

3c)

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  1.5727575  0.6619591  2.375913 2.486134e-02
## oc          -0.2932385  0.1407167 -2.083892 4.676341e-02
## trap         0.2447827  0.1661820  1.472979 1.523197e-01
## voplus       0.8133561  0.1424578  5.709453 4.564402e-06
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    5.2114553  0.4161271 12.523712 3.187292e-13
## logbiomark$loc 0.4404316  0.1227089  3.589238 1.204697e-03
```

```
##                   Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)      5.03717552  0.3855945 13.0634019 1.958962e-13
## logbiomark$loc   0.05674508  0.1847554  0.3071362 7.610124e-01
## logbiomark$ltrap 0.58969044  0.2258420  2.6110756 1.433985e-02
```

Similar to when we applied the log data, the first model overall has lower p-values compared to the rest.

The coefficients in the multiple regression model are more normalized compared to the rest.

3d)

```
##  R-squared Adjusted R-squared Estimate std
##     0.7477             0.7196       0.2559

##  R-squared Adjusted R-squared Estimate std
##     0.3076             0.2837        0.409

##  R-squared Adjusted R-squared Estimate std
##     0.4432             0.4034       0.3733
```

The R^2 indicates the multiple regression model is a good fit between the rest of the models.

3e)

The results from part b, shows when applying logarithm to the data, that there still is a very low p-value for voplus, thus there is strong evidence that a model between voplus and vominus will result in an effective model.

```
##
## Call:
## lm(formula = vominus ~ voplus, data = log(df))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.77773 -0.07606  0.00449  0.16131  0.35885
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.75657    0.59361   2.959  0.00609 **
## voplus       0.73050    0.08776   8.324 3.56e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.267 on 29 degrees of freedom
## Multiple R-squared:  0.7049, Adjusted R-squared:  0.6948
## F-statistic: 69.28 on 1 and 29 DF,  p-value: 3.556e-09

##             Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 1.7565695 0.59361216 2.959120 6.088128e-03
## voplus      0.7304955 0.08776084 8.323708 3.556275e-09

##  R-squared Adjusted R-squared Estimate std
##     0.7049             0.6948        0.267
```

In comparison with the multiple regression model, a model based on voplus and vominus has a much lower p-value for vominus, although the p-value for intercept is a power of ten higher. Further proving there is a strong relation between voplus and vominus.