Exploring Weights Dependence with Body Dimensions Measurements

THROUGH MULTIPLE REGRESSION ANALYSIS

RYAN Z & PRADIP D.

Motivation and project outline

- To design a regression model for predicting or estimating weight dependence using body dimension measurement
- Separated data into training and test set
- Adequately check regression models through internal analysis
  - Residual analysis
  - Variable inflation factors
  - High Leverage
  - Outliers
- Model validation
- Testing data set

Key assumptions for linear regression analysis

▶ The response variable y, must be independent and identically collected and follow normal distribution

▶ The error term ϵ, of the model, must have an E(ϵ) = 0 and Var(ϵ) = σ^2.

▶ The errors are constant and uncorrelated.

▶ The regressors are random variables.

▶ Linearity of the model.

Exploration of data

- ► **Dimension of original data set:**
  - 25 variables
  - 507 observations.
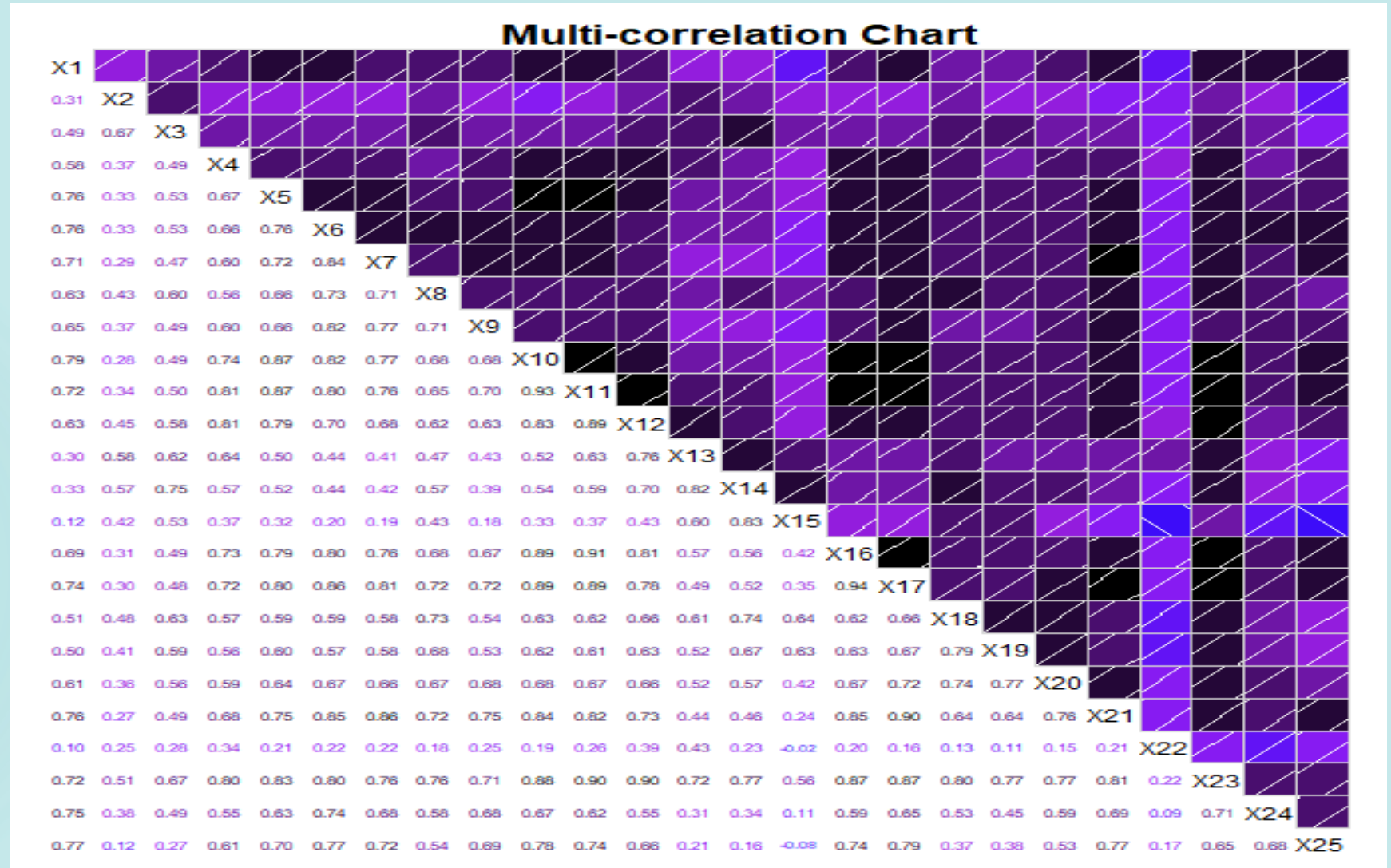- ► **Variables:**
  - 24 continuous
  - 1 discreet

| Variable Names	Units	Type	Min	1st Qu	Median	Mean	3rd Qu	Max
Biacromial diameter	cm	Continuous	32.4	36.2	38.7	38.82	41.2	47.4
Biiliac diameter	cm	Continuous	19.4	26.5	28	27.88	29.4	34.7
Bitrochanteric diameter	cm	Continuous	24.7	30.6	32	31.99	33.5	38
Chest depth	cm	Continuous	14.3	17.3	19	19.24	20.9	27.5
Chest diameter	cm	Continuous	22.3	25.7	27.7	27.99	20	35.6
Elbow diameter	cm	Continuous	9.9	12.4	13.2	13.38	14.4	16.7
Wrist diameter	cm	Continuous	8.1	9.8	10.5	10.54	11.2	13.3
Knee diameter	cm	Continuous	15.7	17.9	18.8	18.82	19.7	24.3
Ankle diameter	cm	Continuous	9.9	13	13.8	13.87	14.8	17.2
Shoulder girth	cm	Continuous	85.9	99.5	108.3	108.2	116.6	134.8
Chest girth	cm	Continuous	72.6	85.3	91.8	93.42	101.2	118.7
Waist girth	cm	Continuous	57.9	68	76	77.15	84.9	113.2
Navel girth	cm	Continuous	64	78.8	85	85.86	92	121.1
Hip girth	cm	Continuous	80.7	92	96	96.78	101	128.3
Thigh girth	cm	Continuous	46.8	53.7	56.3	56.9	59.5	75.7
Bicep girth	cm	Continuous	22.4	27.6	31	31.2	34.5	42.4
Forearm girth	cm	Continuous	19.6	23.6	25.8	25.97	28.4	32.5
Knee girth	cm	Continuous	29	34.5	36	36.25	38	49
Calf max girth	cm	Continuous	28.4	34.1	35.9	36.12	38	47.7
Ankle min girth	cm	Continuous	17.4	21	22	22.21	23.3	29.3
Wrist min girth	cm	Continuous	13	15	16.1	16.11	17.1	19.6
Age	years	Continuous	18	23	27	30.36	36	67
Weight	kg	Continuous	42	58.2	68.2	69.28	79.1	116.4
Height	cm	Continuous	147.2	164	170.2	171.1	177.8	198.1
Gender	1-M, 0-F	Discrete	0	0	0	0.4858	1	1

Exploring the data
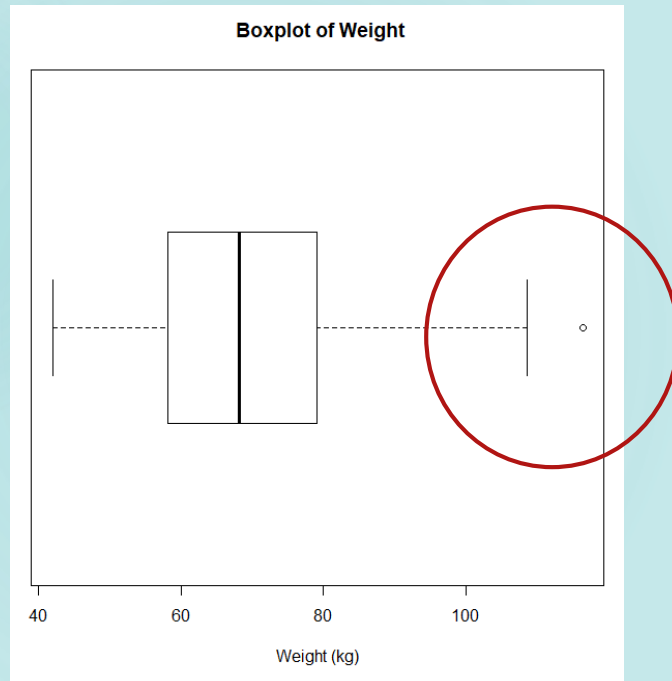
- ▶ **Correlation of variables**

- ▶ **X23 is response variable**

- ▶ **Darker color = greater correlation value**

- ▶ **Predicting muticollinearity**

Multi-correlation Chart

Checking the normality of the response variable

- ▶ **Outliers present**
- ▶ **Right skewed**
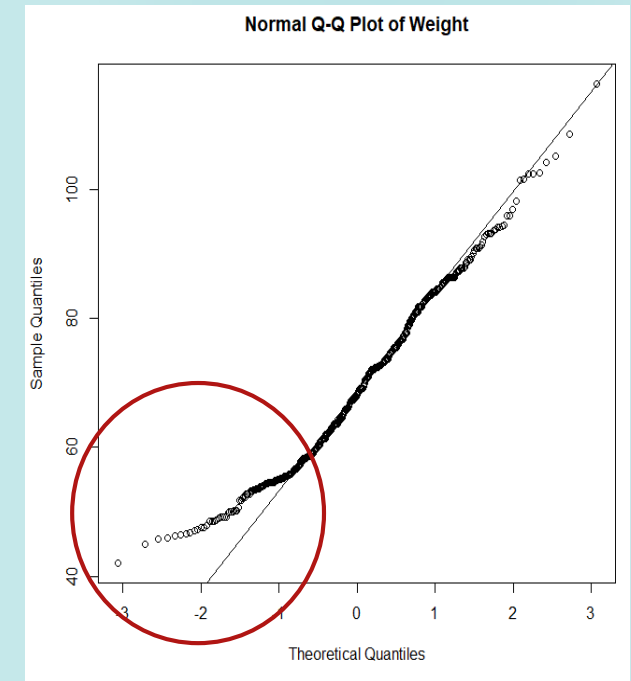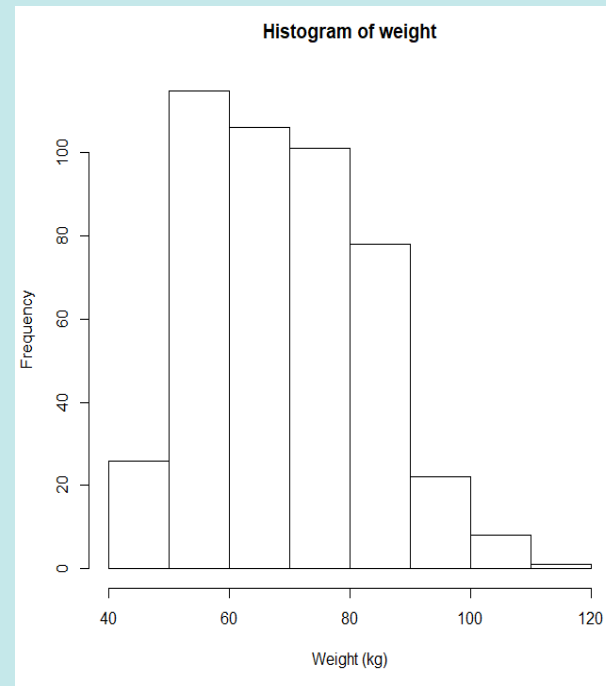
Original regression model Summary & ANOVA tables

► **Weight (response variable) vs. 24 (regressor variables)**

```
Residuals:
    Min      1Q   Median      3Q      Max
-7.9425 -1.3865   0.0686  1.3295   8.7887

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -122.29378    2.85120 -42.892  < 2e-16 ***
biacromial_diam       -0.05731    0.07152  -0.801 0.423348
biiliac_diam           0.15991    0.06936   2.306 0.021607 *
birtrochan_diam       -0.09478    0.09643  -0.983 0.326207
chest_depth            0.27573    0.07444   3.704 0.000240 ***
chest_diam             0.15096    0.08487   1.779 0.075999 .
elbow_diam             0.20570    0.19823   1.038 0.299989
wrist_diam             0.20769    0.23415   0.887 0.375571
knee_diam              0.45182    0.14268   3.167 0.001651 **
ankle_diam             0.07248    0.16546   0.438 0.661571
shoulder_girth         0.07907    0.03337   2.370 0.018250 *
chest_girth            0.14773    0.04107   3.597 0.000359 ***
waist_girth            0.36776    0.02946  12.484  < 2e-16 ***
navel_girth           -0.00489    0.02646  -0.185 0.853482
hip_girth              0.22496    0.04933   4.561 6.65e-06 ***
thigh_girth            0.25732    0.05673   4.536 7.44e-06 ***
bicep_girth            0.05639    0.08950   0.630 0.529025
forearm_girth          0.48394    0.14713   3.289 0.001087 **
knee_girth             0.18893    0.08266   2.286 0.022756 *
calf_max_girth         0.36910    0.07132   5.175 3.50e-07 ***
ankle_min_girth       -0.03030    0.10853  -0.279 0.780203
wrist_min_girth       -0.18119    0.22077  -0.821 0.412252
age                   -0.05720    0.01315  -4.351 1.69e-05 ***
height                 0.31064    0.01953  15.906  < 2e-16 ***
as.factor(gender)1    -1.36287    0.55059  -2.475 0.013695 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.116 on 432 degrees of freedom
Multiple R-squared:  0.9767,   Adjusted R-squared:  0.9754
F-statistic: 754.6 on 24 and 432 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: weight
                   Df Sum Sq Mean Sq   F value     Pr(>F)
biacromial_diam     1  42649   42649 9528.1654  < 2.2e-16 ***
biiliac_diam        1   7353    7353 1642.7625  < 2.2e-16 ***
birtrochan_diam     1   4428    4428  989.1844  < 2.2e-16 ***
chest_depth         1  12256   12256 2738.2125  < 2.2e-16 ***
chest_diam          1   4021    4021  898.2346  < 2.2e-16 ***
elbow_diam          1   1097    1097  245.1117  < 2.2e-16 ***
wrist_diam          1    492     492  109.8425  < 2.2e-16 ***
knee_diam           1    811     811  181.1709  < 2.2e-16 ***
ankle_diam          1     40      40    9.0122   0.002837 **
shoulder_girth      1   2168    2168  484.3372  < 2.2e-16 ***
chest_girth         1    659     659  147.1435  < 2.2e-16 ***
waist_girth         1   1100    1100  245.7582  < 2.2e-16 ***
navel_girth         1    137     137   30.6177  5.457e-08 ***
hip_girth           1   1521    1521  339.7148  < 2.2e-16 ***
thigh_girth         1    474     474  105.8210  < 2.2e-16 ***
bicep_girth         1      7       7    1.6610   0.198157
forearm_girth       1    155     155   34.6062  8.097e-09 ***
knee_girth          1    303     303   67.7927  2.175e-15 ***
calf_max_girth      1     71      71   15.8467  8.056e-05 ***
ankle_min_girth     1      2       2    0.3616   0.547909
wrist_min_girth     1      3       3    0.6916   0.406068
age                 1    188     188   41.9366  2.568e-10 ***
height              1   1106    1106  247.0421  < 2.2e-16 ***
as.factor(gender)   1     27      27    6.1271   0.013695 *
Residuals         432   1934       4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
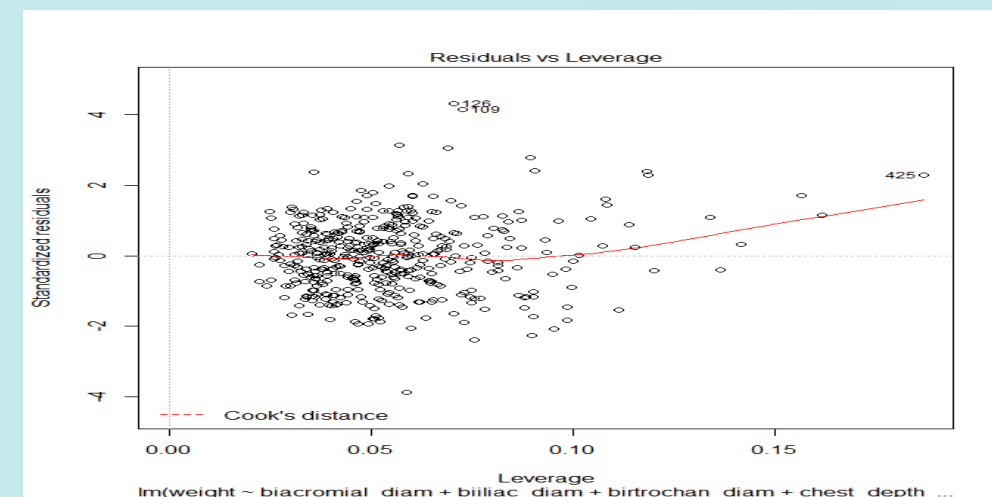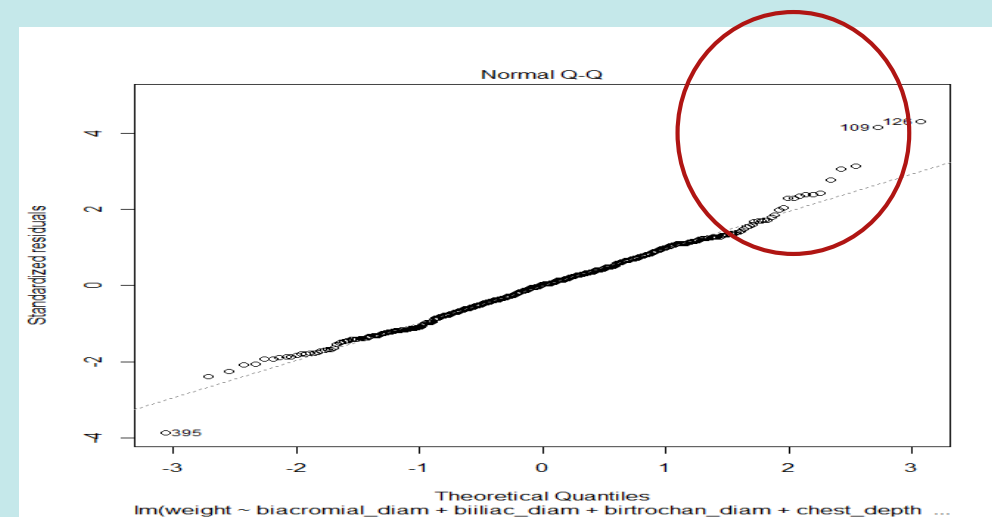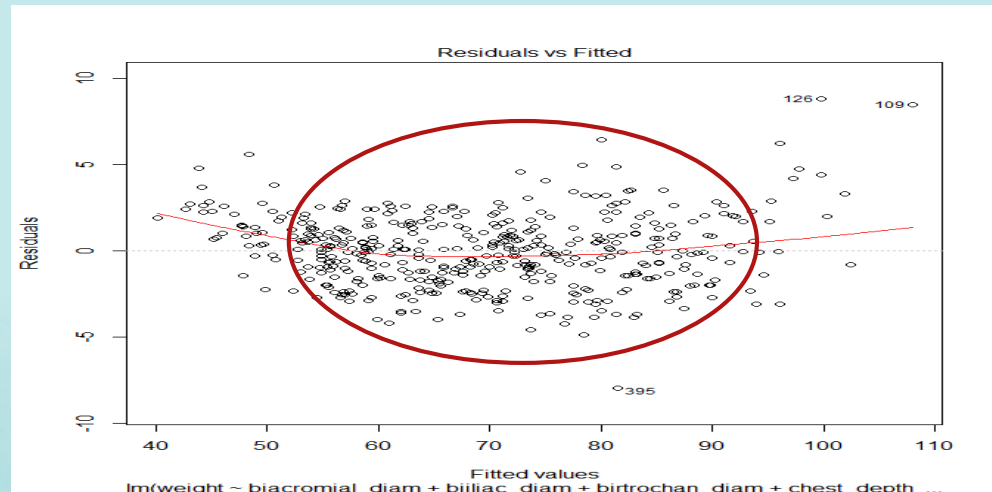
Residual plots error analysis

Partial residual plots

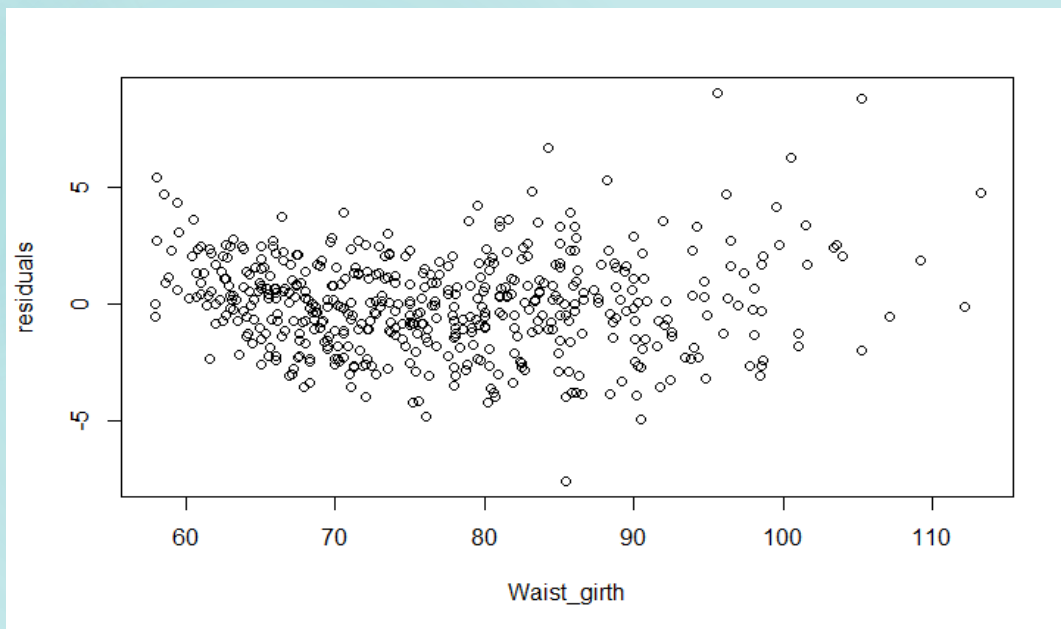- **Partial residual plots show similar pattern of regressor and response variable plots**
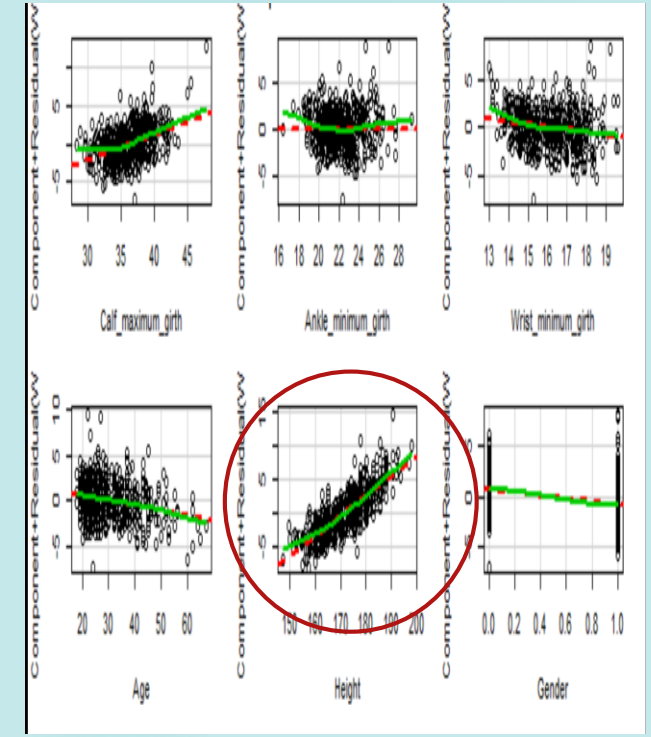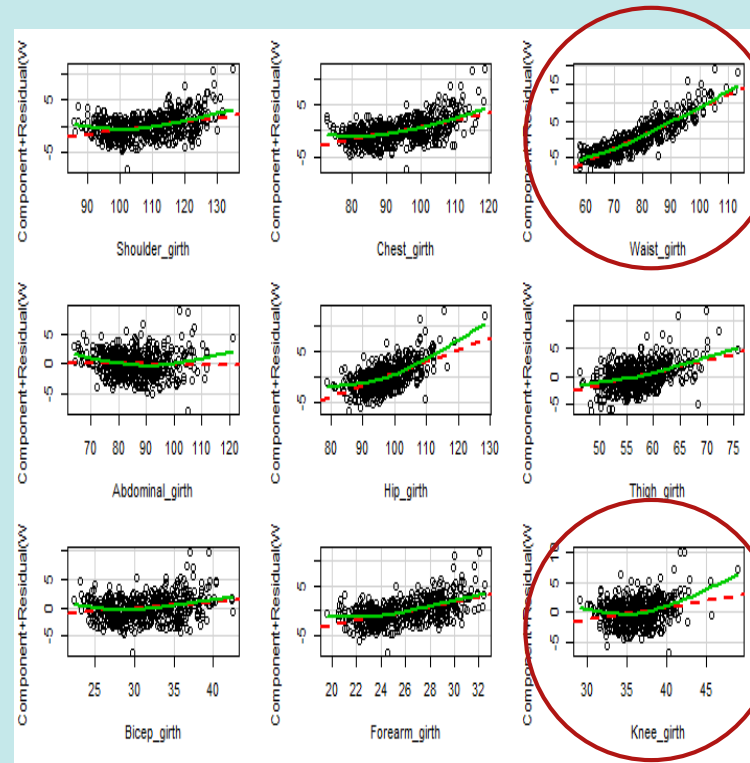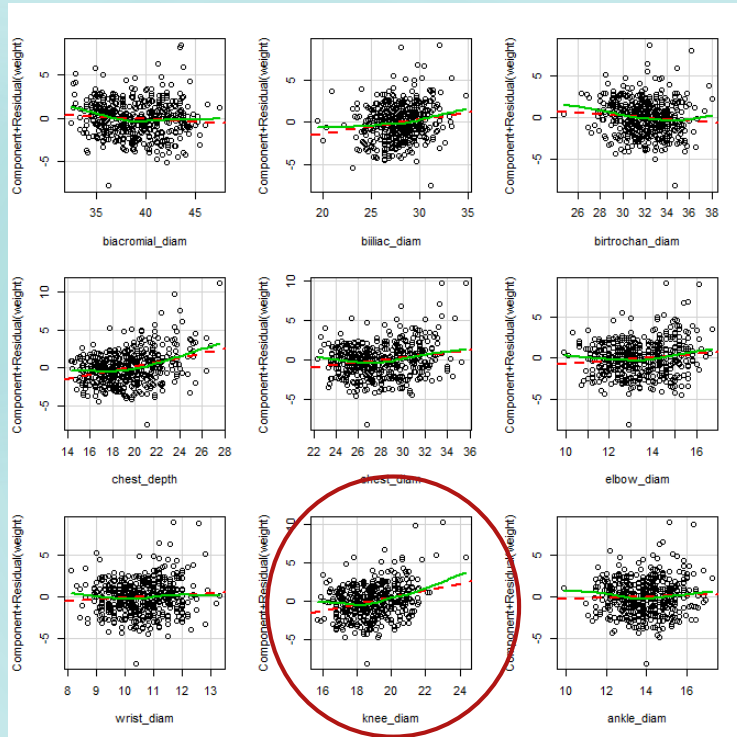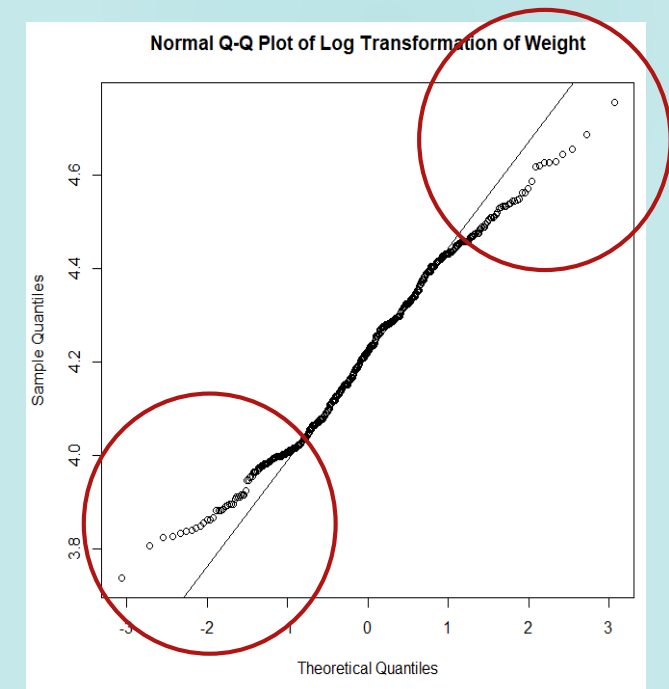
Partial residual plots

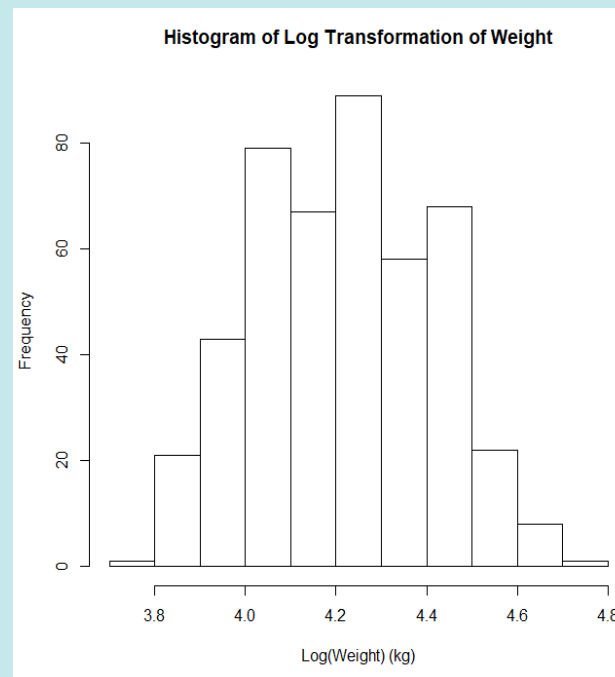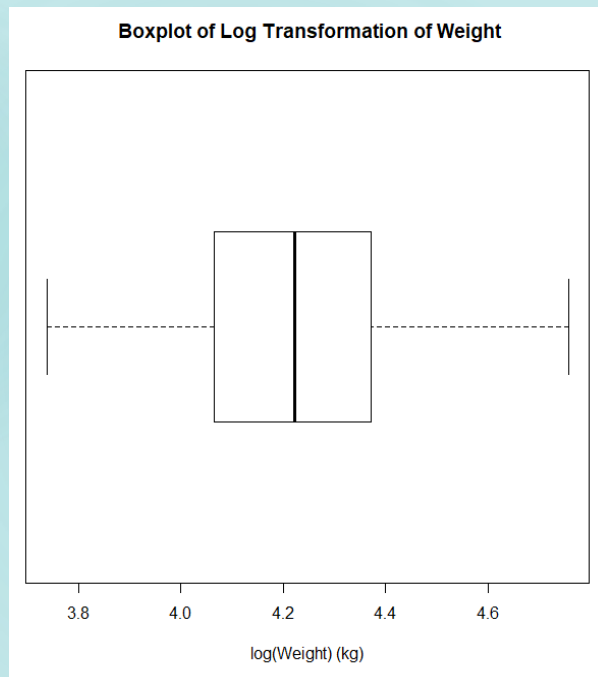- ▶ **The usefulness of partial residual plot.**

- ▶ **The residuals vs. waist girth**

- ▶ **The response variable weight is related to the waist girth indicating the extent and direction of linearity.**
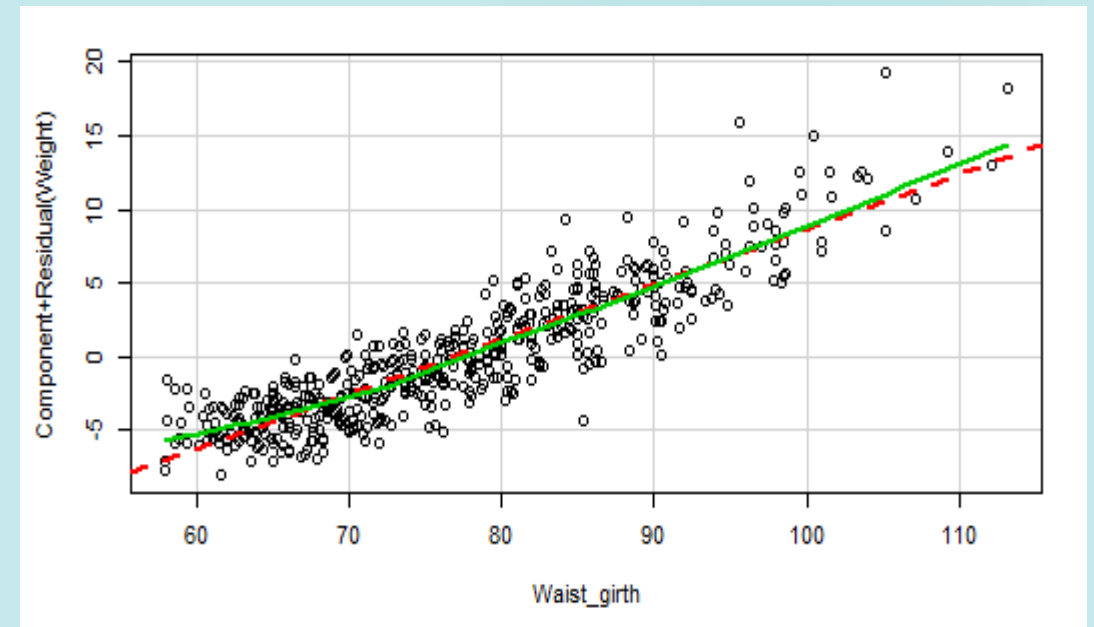
Log transformation of response variable

- ▶ **No outliers**
- ▶ **Skewed tails**

Log transformation regression model summary & ANOVA tables

```
Residuals:
    Min      1Q   Median      3Q     Max
-4.1712 -1.3257  0.0946  1.2491  5.3429

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -1.195e+02  2.583e+00 -46.278  < 2e-16 ***
biacromial_diam      -2.651e-03  6.437e-02  -0.041 0.967172
biiliac_diam          1.821e-01  6.594e-02   2.761 0.006017 **
birtrochan_diam      -6.670e-02  8.572e-02  -0.778 0.436915
chest_depth           2.372e-01  6.719e-02   3.531 0.000461 ***
chest_diam            1.481e-01  7.704e-02   1.923 0.055232 .
elbow_diam            2.262e-01  1.791e-01   1.263 0.207288
wrist_diam            1.901e-01  2.099e-01   0.906 0.365585
knee_diam             2.003e-01  1.391e-01   1.441 0.150455
ankle_diam            1.771e-01  1.516e-01   1.168 0.243546
shoulder_girth        3.002e-02  3.007e-02   0.998 0.318810
chest_girth           1.481e-01  3.706e-02   3.997 7.60e-05 ***
waist_girth           3.712e-01  2.734e-02  13.576  < 2e-16 ***
navel_girth          -2.315e-02  2.353e-02  -0.984 0.325850
hip_girth             2.130e-01  4.419e-02   4.821 2.02e-06 ***
thigh_girth           2.936e-01  5.253e-02   5.590 4.14e-08 ***
bicep_girth           1.274e-01  8.251e-02   1.544 0.123418
forearm_girth         4.606e-01  1.305e-01   3.529 0.000464 ***
knee_girth            1.983e-01  7.671e-02   2.584 0.010101 *
calf_max_girth        3.402e-01  6.544e-02   5.199 3.16e-07 ***
ankle_min_girth       4.648e-04  9.688e-02   0.005 0.996175
wrist_min_girth      -1.875e-01  1.982e-01  -0.946 0.344599
age                  -4.613e-02  1.222e-02  -3.775 0.000184 ***
height                3.174e-01  1.748e-02  18.163  < 2e-16 ***
as.factor(gender)1   -1.077e+00  5.092e-01  -2.115 0.034995 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.837 on 411 degrees of freedom
Multiple R-squared:  0.9806,   Adjusted R-squared:  0.9795
F-statistic: 866.6 on 24 and 411 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: weight
                   Df Sum Sq Mean Sq   F value    Pr(>F)
biacromial_diam     1  38727   38727 11476.5092 < 2.2e-16 ***
biiliac_diam        1   5432    5432  1609.7539 < 2.2e-16 ***
birtrochan_diam     1   3362    3362   996.2804 < 2.2e-16 ***
chest_depth         1  10603   10603  3142.2221 < 2.2e-16 ***
chest_diam          1   3231    3231   957.5318 < 2.2e-16 ***
elbow_diam          1   1182    1182   350.3379 < 2.2e-16 ***
wrist_diam          1    342     342   101.4765 < 2.2e-16 ***
knee_diam           1    387     387   114.7345 < 2.2e-16 ***
ankle_diam          1      1       1     0.3898   0.53276
shoulder_girth      1   1628    1628   482.5648 < 2.2e-16 ***
chest_girth         1    707     707   209.5473 < 2.2e-16 ***
waist_girth         1    949     949   281.2124 < 2.2e-16 ***
navel_girth         1     69      69    20.5593 7.594e-06 ***
hip_girth           1   1338    1338   396.4713 < 2.2e-16 ***
thigh_girth         1    461     461   136.5416 < 2.2e-16 ***
bicep_girth         1     21      21     6.2684   0.01268 *
forearm_girth       1    133     133    39.5215 8.295e-10 ***
knee_girth          1    289     289    85.6965 < 2.2e-16 ***
calf_max_girth      1     75      75    22.2433 3.296e-06 ***
ankle_min_girth     1      1       1     0.2267   0.63424
wrist_min_girth     1      1       1     0.3375   0.56162
age                 1    120     120    35.4714 5.561e-09 ***
height              1   1104    1104   327.1693 < 2.2e-16 ***
as.factor(gender)   1     15      15     4.4750   0.03499 *
Residuals         411   1387       3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variable selection

- **Forward, backward, and stepwise variable selection.**

- **The summaries are all the same**

```
Forward:
Step:  AIC=765.15
weight ~ waist_girth + knee_girth + height + thigh_girth + forearm_girth +
    chest_girth + calf_max_girth + hip_girth + knee_diam + age +
    chest_depth + shoulder_girth + as.factor(gender) + biiliac_diam +
    chest_diam + elbow_diam

Summary:
Residual standard error: 2.092 on 490 degrees of freedom
Multiple R-squared:  0.9762,   Adjusted R-squared:  0.9754
F-statistic:  1256 on 16 and 490 DF,  p-value: < 2.2e-16


Backward:
Step:  AIC=765.15
weight ~ biiliac_diam + chest_depth + chest_diam + elbow_diam +
    knee_diam + shoulder_girth + chest_girth + waist_girth +
    hip_girth + thigh_girth + forearm_girth + knee_girth + calf_max_girth +
    age + height + as.factor(gender)

Summary:
Residual standard error: 2.092 on 490 degrees of freedom
Multiple R-squared:  0.9762,   Adjusted R-squared:  0.9754
F-statistic:  1256 on 16 and 490 DF,  p-value: < 2.2e-16


Stepwise:
Step:  AIC=765.15
weight ~ waist_girth + knee_girth + height + thigh_girth + forearm_girth +
    chest_girth + calf_max_girth + hip_girth + knee_diam + age +
    chest_depth + shoulder_girth + as.factor(gender) + biiliac_diam +
    chest_diam + elbow_diam

Summary:
Residual standard error: 2.092 on 490 degrees of freedom
Multiple R-squared:  0.9762,   Adjusted R-squared:  0.9754
F-statistic:  1256 on 16 and 490 DF,  p-value: < 2.2e-16
```

Stepwise selection summary & ANOVA tables

```
Residuals:
    Min      1Q  Median      3Q     Max
-7.8829 -1.4105  0.0389  1.2331  8.6502

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -123.70755    2.66926 -46.345  < 2e-16 ***
waist_girth           0.36518    0.02683  13.609  < 2e-16 ***
knee_girth            0.17369    0.07832   2.218 0.027079 *
height                0.31013    0.01776  17.462  < 2e-16 ***
thigh_girth           0.27212    0.05101   5.334 1.54e-07 ***
chest_girth           0.17140    0.03775   4.541 7.24e-06 ***
calf_max_girth        0.34841    0.06398   5.445 8.60e-08 ***
forearm_girth         0.54927    0.10264   5.351 1.41e-07 ***
hip_girth             0.20330    0.04036   5.037 6.92e-07 ***
knee_diam             0.49951    0.12818   3.897 0.000113 ***
age                  -0.05536    0.01241  -4.462 1.03e-05 ***
chest_depth           0.26463    0.07315   3.618 0.000332 ***
as.factor(gender)1   -1.32321    0.50490  -2.621 0.009078 **
shoulder_girth        0.07422    0.03112   2.385 0.017481 *
biiliac_diam          0.13694    0.06066   2.258 0.024462 *
chest_diam            0.12196    0.07970   1.530 0.126701
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.109 on 441 degrees of freedom
Multiple R-squared:  0.9764,   Adjusted R-squared:  0.9756
F-statistic:  1215 on 15 and 441 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: weight
                   Df Sum Sq Mean Sq   F value    Pr(>F)
waist_girth         1  67781   67781 15241.2261 < 2.2e-16 ***
knee_girth          1   6168    6168  1386.9699 < 2.2e-16 ***
height              1   3166    3166   711.9122 < 2.2e-16 ***
thigh_girth         1   1804    1804   405.6006 < 2.2e-16 ***
chest_girth         1   1297    1297   291.6468 < 2.2e-16 ***
calf_max_girth      1    268     268    60.1955 6.025e-14 ***
forearm_girth       1    135     135    30.3754 6.067e-08 ***
hip_girth           1    164     164    36.8573 2.743e-09 ***
knee_diam           1     61      61    13.8228 0.0002268 ***
age                 1     64      64    14.4493 0.0001642 ***
chest_depth         1     43      43     9.7541 0.0019071 **
as.factor(gender)   1     26      26     5.8192 0.0162590 *
shoulder_girth      1     27      27     6.1514 0.0135034 *
biiliac_diam        1     24      24     5.4310 0.0202331 *
chest_diam          1     10      10     2.3413 0.1267015
Residuals         441   1961       4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All possible regressors

Number of regressors	r_squared	rss	r_adj	MSres	Mallow CP	bic
1	0.81663	15220	0.81623	30.13807	2947.25	-762.94
2	0.904463	7929.6	0.90404	15.7333	1320.55	-1054.8
3	0.944875	4575.4	0.94451	9.09626	573.197	-1300
4	0.962914	3078.1	0.96259	6.131745	240.688	-1475
5	0.967814	2671.4	0.96746	5.332213	151.828	-1533.6
6	0.970367	2459.5	0.96997	4.919054	106.485	-1565.3
7	0.972481	2284.1	0.97205	4.577311	69.2874	-1592.9
8	0.973627	2189	0.97316	4.39559	50.0468	-1606.3
9	0.974435	2121.9	0.97392	4.269496	37.0639	-1614.3
10	0.974955	2078.8	0.97439	4.19104	29.4163	-1617.6
11	0.975365	2044.7	0.97476	4.130746	23.8121	-1619
12	0.97571	2016.1	0.97505	4.081156	19.4162	-1619.4
13	0.975993	1992.6	0.97529	4.041784	16.168	-1618.6
14	0.976246	1971.6	0.97549	4.007361	13.4813	-1617.3
15	0.976371	1961.2	0.97557	3.994317	13.1551	-1613.6
16	**0.976458**	**1954**	**0.9756**	**3.987771**	**13.5462**	**-1609.2**
17	0.976545	1946.7	0.97564	3.981059	13.922	-1604.7
18	0.9766	1942.2	0.97564	3.979927	14.9091	-1599.7
19	0.976629	1939.8	0.97561	3.983198	16.3759	-1594.1
20	0.97667	1936.4	0.9756	3.984284	17.6039	-1588.8
21	0.97669	1934.8	0.97556	3.989199	19.2463	-1583.1
22	0.976696	1934.2	0.97552	3.9963	21.1229	-1577.1
23	0.976701	1933.8	0.97546	4.003751	23.0341	-1571
24	0.976703	1933.7	0.97541	4.01174	25	-1564.9

Influence measures of original regression model

- ► **Selection Criteria for influential points**
  - ➤ **n=557, p=25**
  - ➤ **$|DFBETAS_{j,i}| > 2/sqrt(n) = 0.08882312$**
  - ➤ **$DFFITS_i > 2*sqrt(p/n) = 0.4351426$**
  - ➤ **$D_i > 1$**
  - ➤ **$COVARATIO_i < 1$**
- ► **Took out all of the outliers with the ***
- ► **Checked normality of response variable then ran a new regression model**
- ► **New dimensions are 536 observations and 25 variables**
- ► **Compared the summary graphs of the initial and new regression models**

Normality of response variable after outliers taken out

- ▶ **Normal**
- ▶ **Skewed tails**
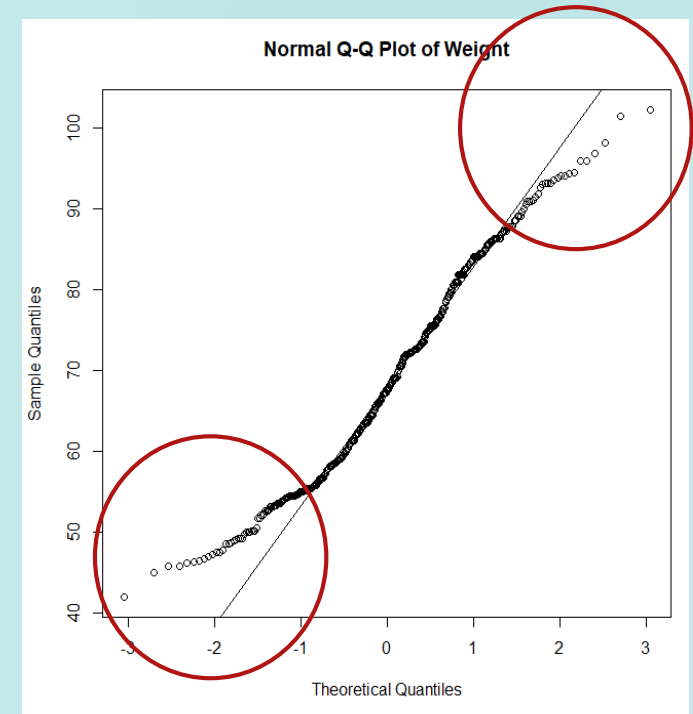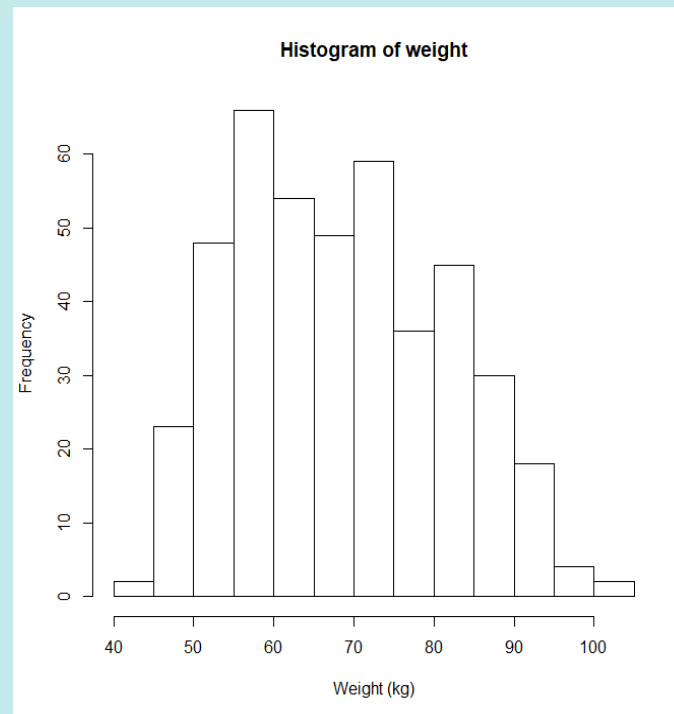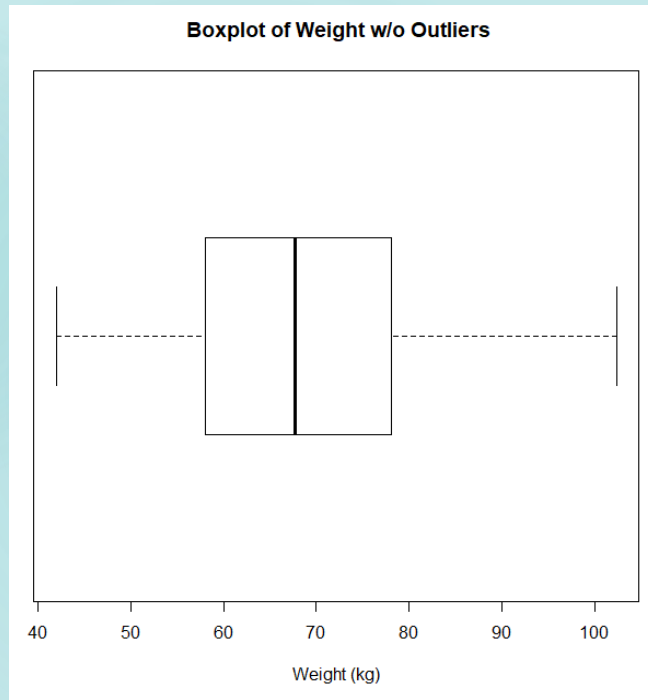
New model summary & ANOVA tables after outliers removed

```
Residuals:
     Min        1Q    Median        3Q       Max
-0.093001 -0.019602 -0.000356  0.018135  0.085731

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.4180744  0.0390687  36.297  < 2e-16 ***
biacromial_diam      0.0009672  0.0009800   0.987 0.324232
biiliac_diam         0.0007809  0.0009504   0.822 0.411736
birtrochan_diam     -0.0003111  0.0013213  -0.235 0.813977
chest_depth          0.0018629  0.0010200   1.826 0.068480 .
chest_diam           0.0002101  0.0011630   0.181 0.856741
elbow_diam           0.0003207  0.0027162   0.118 0.906064
wrist_diam           0.0039604  0.0032085   1.234 0.217743
knee_diam            0.0068510  0.0019551   3.504 0.000506 ***
ankle_diam           0.0031541  0.0022672   1.391 0.164886
shoulder_girth       0.0009634  0.0004572   2.107 0.035701 *
chest_girth          0.0025715  0.0005628   4.569 6.40e-06 ***
waist_girth          0.0040231  0.0004037   9.967  < 2e-16 ***
navel_girth          0.0006884  0.0003626   1.898 0.058308 .
hip_girth            0.0032516  0.0006759   4.811 2.08e-06 ***
thigh_girth          0.0044113  0.0007773   5.675 2.55e-08 ***
bicep_girth          0.0010539  0.0012264   0.859 0.390630
forearm_girth        0.0077562  0.0020160   3.847 0.000137 ***
knee_girth           0.0024861  0.0011326   2.195 0.028692 *
calf_max_girth       0.0055794  0.0009773   5.709 2.12e-08 ***
ankle_min_girth     -0.0019101  0.0014872  -1.284 0.199699
wrist_min_girth     -0.0021771  0.0030251  -0.720 0.472115
age                 -0.0005301  0.0001801  -2.943 0.003423 **
height               0.0045668  0.0002676  17.066  < 2e-16 ***
as.factor(gender)1  -0.0038074  0.0075445  -0.505 0.614060
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02899 on 432 degrees of freedom
Multiple R-squared:  0.9787,   Adjusted R-squared:  0.9776
F-statistic:   829 on 24 and 432 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: transy
                    Df  Sum Sq Mean Sq   F value    Pr(>F)
biacromial_diam      1  9.0850  9.0850 10809.9904 < 2.2e-16 ***
biiliac_diam         1  1.4260  1.4260  1696.8005 < 2.2e-16 ***
birtrochan_diam      1  0.9445  0.9445  1123.8642 < 2.2e-16 ***
chest_depth          1  2.3076  2.3076  2745.8029 < 2.2e-16 ***
chest_diam           1  0.7429  0.7429   884.0068 < 2.2e-16 ***
elbow_diam           1  0.2664  0.2664   316.9405 < 2.2e-16 ***
wrist_diam           1  0.1176  0.1176   139.8923 < 2.2e-16 ***
knee_diam            1  0.1817  0.1817   216.1615 < 2.2e-16 ***
ankle_diam           1  0.0031  0.0031     3.6905 0.055382 .
shoulder_girth       1  0.4631  0.4631   550.9970 < 2.2e-16 ***
chest_girth          1  0.1589  0.1589   189.1052 < 2.2e-16 ***
waist_girth          1  0.1819  0.1819   216.4874 < 2.2e-16 ***
navel_girth          1  0.0495  0.0495    58.8908 1.120e-13 ***
hip_girth            1  0.3047  0.3047   362.5718 < 2.2e-16 ***
thigh_girth          1  0.1041  0.1041   123.9061 < 2.2e-16 ***
bicep_girth          1  0.0060  0.0060     7.1944 0.007594 **
forearm_girth        1  0.0363  0.0363    43.2099 1.420e-10 ***
knee_girth           1  0.0531  0.0531    63.2139 1.635e-14 ***
calf_max_girth       1  0.0130  0.0130    15.4427 9.901e-05 ***
ankle_min_girth      1  0.0002  0.0002     0.2275 0.633597
wrist_min_girth      1  0.0003  0.0003     0.3670 0.544960
age                  1  0.0236  0.0236    28.1037 1.838e-07 ***
height               1  0.2503  0.2503   297.8406 < 2.2e-16 ***
as.factor(gender)    1  0.0002  0.0002     0.2547 0.614060
Residuals          432  0.3631  0.0008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

New model summary and ANOVA w/ log transformation & outliers removed

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.093854 -0.019284 -0.000504  0.017682  0.084670

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.4089455  0.0389059  36.214  < 2e-16 ***
biacromial_diam      0.0008477  0.0009767   0.868 0.385934
biiliac_diam         0.0009466  0.0009479   0.999 0.318534
birtrochan_diam     -0.0005199  0.0013216  -0.393 0.694212
chest_depth          0.0014151  0.0010271   1.378 0.168998
chest_diam           0.0002256  0.0011771   0.192 0.848089
elbow_diam           0.0005179  0.0027220   0.190 0.849182
wrist_diam           0.0033730  0.0032054   1.052 0.293276
knee_diam            0.0071327  0.0019496   3.659 0.000286 ***
ankle_diam           0.0036157  0.0022853   1.582 0.114357
shoulder_girth       0.0009855  0.0004571   2.156 0.031630 *
chest_girth          0.0025453  0.0005658   4.499 8.86e-06 ***
waist_girth          0.0040967  0.0004023  10.182  < 2e-16 ***
navel_girth          0.0006752  0.0003611   1.870 0.062229 .
hip_girth            0.0032355  0.0006754   4.791 2.31e-06 ***
thigh_girth          0.0046783  0.0007772   6.019 3.81e-09 ***
bicep_girth          0.0005922  0.0012232   0.484 0.628556
forearm_girth        0.0079360  0.0020224   3.924 0.000102 ***
knee_girth           0.0023485  0.0011464   2.049 0.041128 *
calf_max_girth       0.0054859  0.0009785   5.607 3.74e-08 ***
ankle_min_girth     -0.0020388  0.0014832  -1.375 0.169990
wrist_min_girth     -0.0009500  0.0030558  -0.311 0.756052
age                 -0.0005533  0.0001791  -3.088 0.002146 **
height               0.0045620  0.0002686  16.987  < 2e-16 ***
as.factor(gender)1  -0.0036824  0.0075357  -0.489 0.625334
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02867 on 421 degrees of freedom
Multiple R-squared:  0.9793,   Adjusted R-squared:  0.9782
F-statistic: 831.5 on 24 and 421 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: newtransy
                    Df  Sum Sq Mean Sq   F value    Pr(>F)
biacromial_diam       1  8.9671  8.9671 10909.4566 < 2.2e-16 ***
biiliac_diam          1  1.3685  1.3685  1664.9381 < 2.2e-16 ***
birtrochan_diam       1  0.9016  0.9016  1096.9294 < 2.2e-16 ***
chest_depth           1  2.2293  2.2293  2712.1402 < 2.2e-16 ***
chest_diam            1  0.7491  0.7491   911.3669 < 2.2e-16 ***
elbow_diam            1  0.2748  0.2748   334.3296 < 2.2e-16 ***
wrist_diam            1  0.1055  0.1055   128.3338 < 2.2e-16 ***
knee_diam             1  0.1787  0.1787   217.4459 < 2.2e-16 ***
ankle_diam            1  0.0035  0.0035     4.2695 0.0394136 *
shoulder_girth        1  0.4539  0.4539   552.2165 < 2.2e-16 ***
chest_girth           1  0.1585  0.1585   192.8661 < 2.2e-16 ***
waist_girth           1  0.1797  0.1797   218.6661 < 2.2e-16 ***
navel_girth           1  0.0535  0.0535    65.0601 7.638e-15 ***
hip_girth             1  0.3018  0.3018   367.1309 < 2.2e-16 ***
thigh_girth           1  0.1040  0.1040   126.5300 < 2.2e-16 ***
bicep_girth           1  0.0050  0.0050     6.1411 0.0135988 *
forearm_girth         1  0.0398  0.0398    48.3765 1.351e-11 ***
knee_girth            1  0.0510  0.0510    62.0606 2.867e-14 ***
calf_max_girth        1  0.0121  0.0121    14.6952 0.0001457 ***
ankle_min_girth       1  0.0002  0.0002     0.2385 0.6255128
wrist_min_girth       1  0.0002  0.0002     0.2848 0.5938485
age                   1  0.0245  0.0245    29.8657 7.943e-08 ***
height                1  0.2410  0.2410   293.2472 < 2.2e-16 ***
as.factor(gender)     1  0.0002  0.0002     0.2388 0.6253341
Residuals           421  0.3460  0.0008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
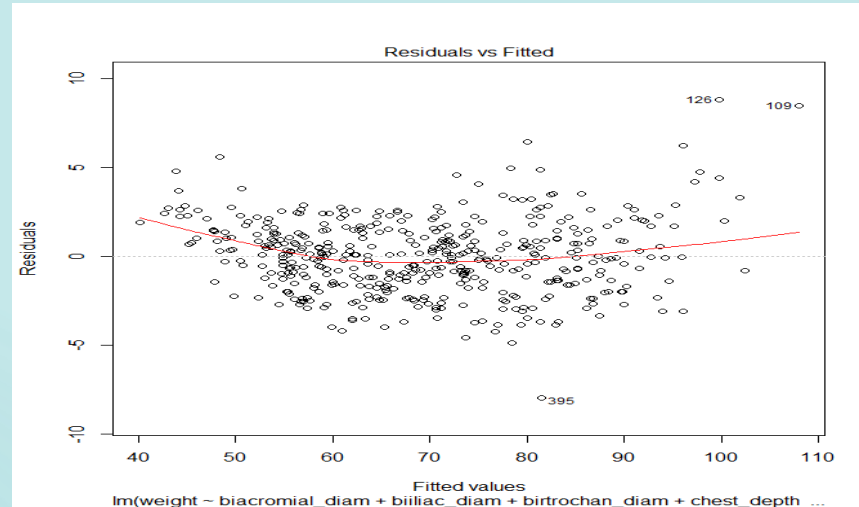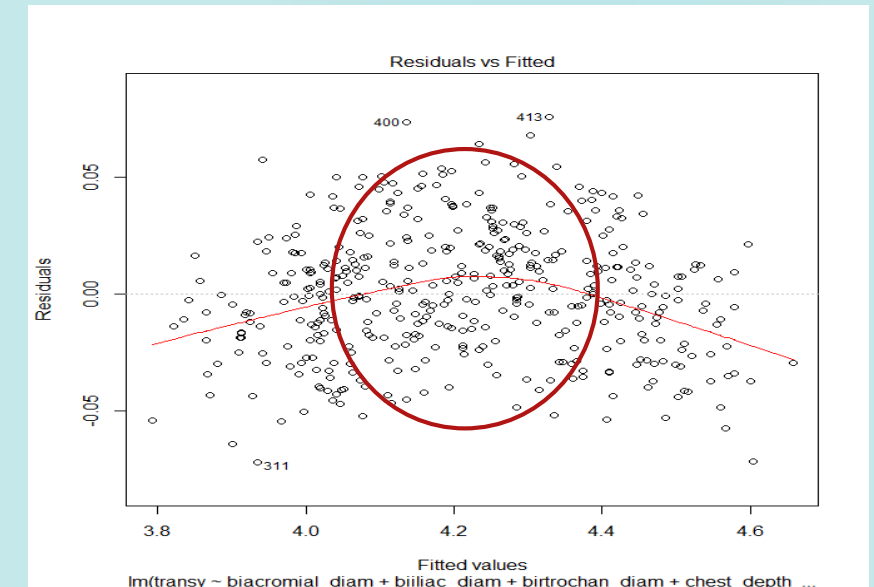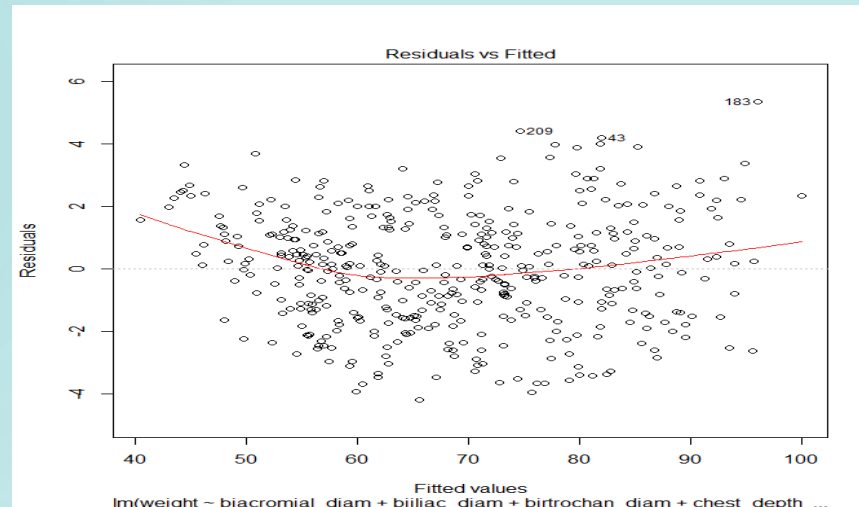
Outlier influence measure plots 1/2
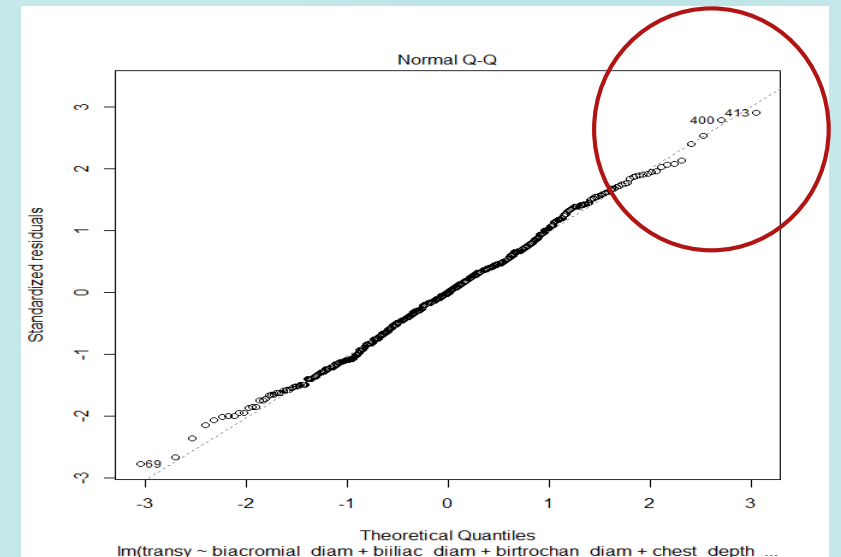
Before removal of outliers

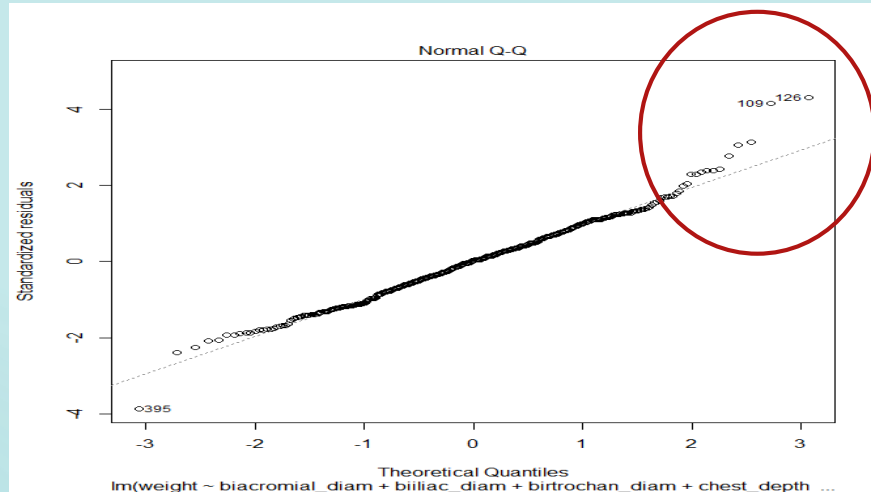**After removal of outliers**

Log transformation

Before removal of outliers
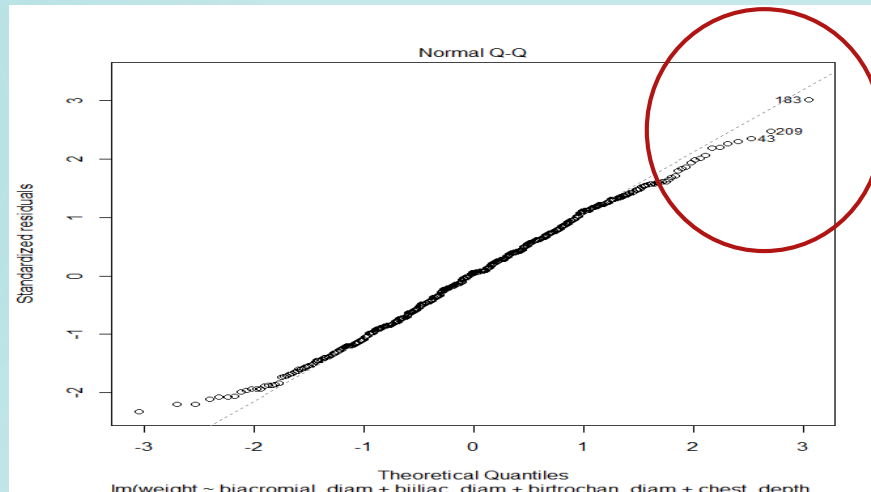
After removal of outliers

Log transformation

Stepwise selection summary & ANOVA tables after outliers removed

```
Residuals:
    Min      1Q   Median      3Q     Max
-4.4591  -1.3545  0.0897  1.2346  5.6960

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -119.06920    2.45649 -48.471  < 2e-16 ***
chest_girth            0.16469    0.03430   4.801 2.20e-06 ***
knee_girth             0.19007    0.07305   2.602 0.009601 **
height                 0.32000    0.01656  19.322  < 2e-16 ***
hip_girth              0.18158    0.03645   4.982 9.23e-07 ***
waist_girth            0.36471    0.02531  14.408  < 2e-16 ***
thigh_girth            0.30906    0.04987   6.197 1.37e-09 ***
forearm_girth          0.45246    0.12184   3.714 0.000232 ***
calf_max_girth         0.34717    0.05868   5.917 6.84e-09 ***
biiliac_diam           0.15366    0.05609   2.739 0.006418 **
age                   -0.04815    0.01162  -4.143 4.14e-05 ***
chest_depth            0.22219    0.06616   3.358 0.000857 ***
knee_diam              0.24027    0.13026   1.845 0.065802 .
chest_diam             0.16630    0.07111   2.339 0.019818 *
elbow_diam             0.30966    0.15838   1.955 0.051235 .
as.factor(gender)1    -0.83738    0.46970  -1.783 0.075342 .
bicep_girth            0.11479    0.07825   1.467 0.143152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.832 on 419 degrees of freedom
Multiple R-squared:  0.9804,   Adjusted R-squared:  0.9796
F-statistic:  1307 on 16 and 419 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: weight
                   Df Sum Sq Mean Sq   F value    Pr(>F)
chest_girth         1  58142   58142 17332.6412 < 2.2e-16 ***
knee_girth          1   6513    6513  1941.4723 < 2.2e-16 ***
height              1   1543    1543   459.8516 < 2.2e-16 ***
hip_girth           1   2217    2217   660.7885 < 2.2e-16 ***
waist_girth         1    678     678   202.1172 < 2.2e-16 ***
thigh_girth         1    458     458   136.5092 < 2.2e-16 ***
forearm_girth       1    307     307    91.4915 < 2.2e-16 ***
calf_max_girth      1    145     145    43.2044 1.468e-10 ***
biiliac_diam        1     34      34    10.0293  0.001653 **
age                 1     35      35    10.3797  0.001374 **
chest_depth         1     26      26     7.8970  0.005184 **
knee_diam           1     23      23     6.9010  0.008930 **
chest_diam          1     16      16     4.7604  0.029677 *
elbow_diam          1     11      11     3.1560  0.076372 .
as.factor(gender)   1      9       9     2.5798  0.108986
bicep_girth         1      7       7     2.1518  0.143152
Residuals         419   1406       3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All possible regressors after outliers removed

Number of Regressors	R_squared	rss	Adjusted R-squared	MSres	Mallow CP	bic
1	0.812407	13425.65	0.811974	27.79636919	3546.58	-717.482
2	0.903406	6912.994	0.90296	14.34231201	1618.609	-1000.8
3	0.95013	3569.118	0.94978	7.42020331	629.6788	-1282.96
4	0.969271	2199.232	0.968985	4.581733959	225.7245	-1488
5	0.972555	1964.174	0.972236	4.100571642	158.0668	-1531.2
6	0.975269	1769.91	0.974924	3.702740498	102.4983	-1570.53
7	0.977473	1612.222	0.977104	3.379919547	57.76865	-1605.14
8	0.978118	1566.061	0.977708	3.290044692	46.08941	-1611.73
9	0.978588	1532.418	0.978136	3.226143345	38.11953	-1615.12
10	0.979074	1497.6	0.978582	3.159492903	29.80137	-1619.06
11	0.979445	1471.109	0.978911	3.110167396	23.95114	-1620.77
12	0.979768	1447.96	0.979194	3.067711198	19.09098	-1621.6
13	0.980015	1430.252	0.9794	3.036629089	15.84354	-1620.89
14	0.980139	1421.404	0.979479	3.024264164	15.22146	-1617.52
15	0.98026	1412.75	0.9795	3.012260322	14.6569	-1614.1
16	**0.980361**	**1405.532**	**0.979611**	**3.003273175**	**14.51783**	**-1610.26**
17	0.980416	1401.612	0.979619	3.001310548	15.35623	-1605.4
18	0.980476	1397.292	0.979633	2.998481071	16.07608	-1600.67
19	0.980518	1394.296	0.979628	2.998487054	17.18833	-1595.53
20	0.980556	1391.592	0.979619	2.999120064	18.38679	-1590.3
21	0.980591	1389.094	0.979606	3.000203135	19.64663	-1585
22	0.980621	1386.918	0.979589	3.001986524	21.00171	-1579.61
23	0.980621	1386.912	0.979539	3.008486105	23.00002	-1573.53
24	0.980621	1386.912	0.979489	3.015026124	25	-1567.45

Variable inflation factors multicollinearity

▶ VIF > 5 or 10 show high correlation with other regressor variables.

▶ Data set are measurements of body dimensions with high multicollinearity

▶ Could eliminate some regressors but 7 have VIF:

  10 > VIF > 5

  and 6 have VIF:

  VIF > 10

Regressor Variable	Original Model	New Model	Transformed Response	Transformed Response Outliers Removed
Biacromial diameter	4.941976467	4.995330475	4.995330475	4.901647496
Biiliac diameter	2.340825153	2.451641383	2.451641383	2.327498975
Bitrochanteric diameter	4.020182284	3.94863176	3.94863176	4.026100934
Chest depth	3.673202729	3.637283267	3.637283267	3.678976036
Chest diameter	5.636756709	5.612554753	5.612554753	5.722814295
Elbow diameter	7.250286819	7.462154704	7.462154704	7.21875086
Wrist diameter	5.069216406	5.041384446	5.041384446	5.079247287
Knee diameter	3.793436641	4.132046466	4.132046466	3.776224452
Ankle diameter	4.327037765	4.646822044	4.646822044	4.336886658
Shoulder girth	12.21790157	11.86677623	11.86677623	12.2577793
Chest girth	17.55681243	17.26623728	17.26623728	17.70273699
Waist girth	11.05179794	11.18507898	11.18507898	10.96772381
Navel girth	6.535567955	6.038630569	6.038630569	6.44751563
Hip girth	11.27110101	10.20419137	10.20419137	11.31851646
Thigh girth	6.539575137	6.192656628	6.192656628	6.529874736
Bicep girth	14.75861646	15.5211888	15.5211888	14.65148893
Forearm girth	17.54650672	17.13217196	17.13217196	17.68194897
Knee girth	4.910723179	4.782153747	4.782153747	4.987222816
Calf max girth	4.292550244	4.135698336	4.135698336	4.345300762
Ankle min girth	4.197469186	4.183576823	4.183576823	4.22078684
Wrist min girth	9.534298333	9.56566901	9.56566901	9.763843359
Age	1.690230476	1.783416781	1.783416781	1.653832677
Height	3.432034698	3.428461572	3.428461572	3.433118395
Gender	7.73155016	8.367954316	8.367954316	7.699453828

Summary

- **Original model: 24 regressors 457 observations**
- **New model outliers removed: 436 observations**
- **Stepwise model: 15 regressors**
- **Stepwise new model : 16 regressor**
- **Log transformation model largest R-squared**
- **Log transformation new model: 446 observations**

Values	Original model	Outliers removed	Stepwise Selection	Stepwise selection, outliers removed	Original model, y log transformation	Original model, y log transformation, outliers removed
R-Square	0.9767	0.9787	0.9764	0.9804	0.9806	0.9793
Adjusted R-Square	0.9754	0.9776	0.9756	0.9796	0.9795	0.9782
Mallow's Cp	25	25	13.55	14.52	25	25
Mean Square residual	4.48	0.00084	4.45	0.0082	3.375	0.0082

Model validation table of coefficients

- No variables had a negative correlation w/ weight

- Many coefficients are close to zero

- Negative signs prove multicollinearity

Regressor Variables	Correlation	Correlation Outliers Removed	Full Model	New Model	Stepwise	Outliers Removed Stepwise	Transformed Response	Transformed Response Outliers Removed
Intercept			-122.3	-119.53	-123.7	-119.07	1.41	1.41
Biacromial diameter	0.72	0.74	-0.057	-0.0027			0.001	0.0008
Biiliac diameter	0.51	0.51	0.16	0.18	0.137	0.154	0.0014	0.001
Bitrochanteric diameter	0.67	0.67	-0.095	-0.067			-0.0004	-0.0005
Chest depth	0.8	0.81	0.276	0.237	0.265	0.222	0.0013	0.0014
Chest diameter	0.83	0.83	0.151	0.148	0.122	0.166	0.0003	0.0002
Elbow diameter	0.8	0.81	0.21	0.226		0.309	0.0005	0.0005
Wrist diameter	0.76	0.76	0.21	0.19			0.005	0.0034
Knee diameter	0.76	0.77	0.452	0.2	0.499	0.24	0.0047	0.0071
Ankle diameter	0.71	0.74	0.0725	0.177			0.0027	0.0036
Shoulder girth	0.88	0.88	0.079	0.03	0.074		0.00082	0.001
Chest girth	0.9	0.9	0.148	0.148	0.171	0.165	0.0025	0.0025
Waist girth	0.9	0.9	0.368	0.371	0.365	0.365	0.0042	0.0041
Navel girth	0.72	0.69	-0.0049	-0.023			0.0006	0.0007
Hip girth	0.77	0.75	0.225	0.213	0.203	0.182	0.0031	0.0032
Thigh girth	0.56	0.54	0.257	0.293	0.272	0.309	0.0048	0.0047
Bicep girth	0.87	0.87	0.056	0.123		0.115	0.00098	0.0006
Forearm girth	0.87	0.87	0.484	0.461	0.549	0.453	0.0081	0.008
Knee girth	0.8	0.81	0.189	0.198	0.174	0.19	0.0029	0.0023
Calf max girth	0.77	0.77	0.369	0.34	0.347	0.347	0.0056	0.0055
Ankle min girth	0.77	0.76	-0.03	0.000046			-0.002	-0.002
Wrist min girth	0.81	0.82	-0.181	-0.187			-0.0014	-0.001
Age	0.22	0.24	-0.057	-0.046	-0.055	-0.048	-0.0005	-0.0006
Height	0.71	0.74	0.311	0.317	0.31	0.32	0.0046	0.0046
Gender	0.65	0.69	-1.36	-1.077	-1.32	-0.837	-0.0035	-0.0037

Conclusions using test data set

- **Test data set: 25 variables 50 observations**

- **Multiplied coefficients of models to data set**

- **Subtracted actual y by predicted y**

- **Squared the difference**

- **Then summed**

	Original Model	New Model	Stepwise Model	Stepwise New Model	Transformed Model	Transformed New Model
Sum Squared error	199.5227	193.1022	202.3834	199.3124	0.034227	0.033774

References

- [1] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

- [2] Larsen, W. A. and S. J. McCleary [1972], "The use of partial residual plot in regression analysis" Technometrics, 14, 781-790.

- [3] Douglas, C. M., Elizabeth, A. P., Vining, G, G., "Introduction to Linear Regression Analysis", Fifth Edition, Wiley A John Wiley and sons, Inc., Publication, ISBN-13: 978-1118780572.

- [4] Grete, H., Roger, J. P., Carter J. K., "Exploring Relationship in Body Dimensions" Journal of Statistics Education Volume 11, Number 2 (2003)

- [5] http://ww2.amstat.org/publications/jse/jse_data_archive.htm