Prediction of Body Fat Percentage

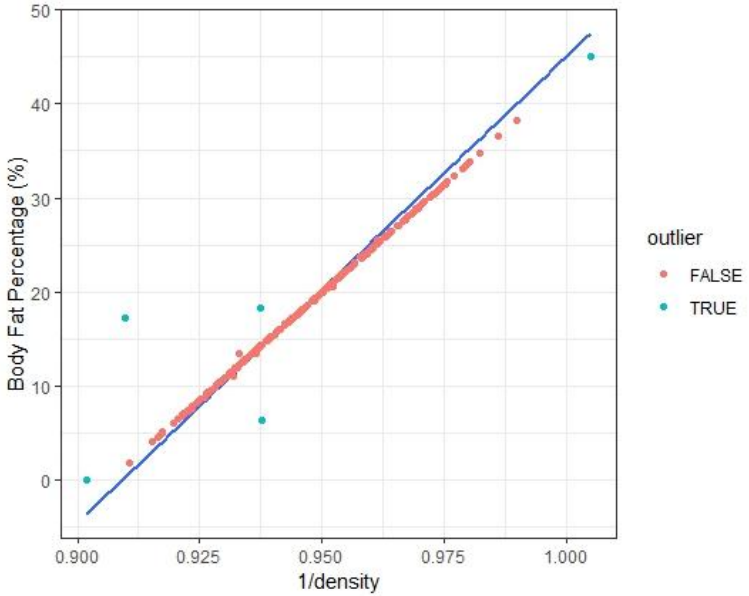
STAT 628 Module 2

Group 1
Yuchen Zeng
Ruixuan Zhao
Jiantong Wang
Hao Pan

Step 1: Raw Data Analysing and Cleaning

Method 1: The relationship between Body fat and density

Percentage of Body Fat (%) = 495/DENSITY - 450



IDNO	BODYFAT from raw data	BODYFAT from formula
96	17.3	0.4
48	6.4	14.1
76	18.3	14.1
182	0.0	-3.6
216	45.1	47.5

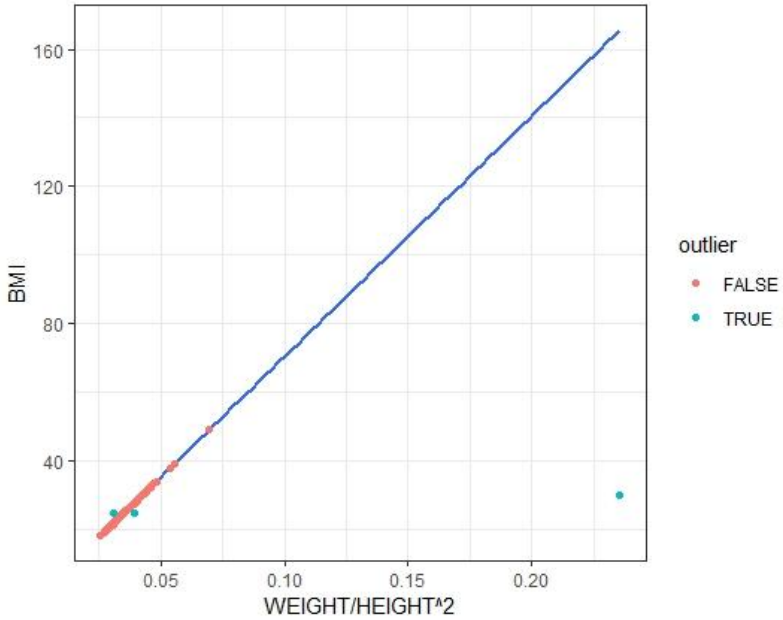
Dealing with the outliers:

We compare the predicted percentage of body fat and the input percentage of body fat and check which one is more reasonable.

- **IDNO-96:** We keep the original percentage of body fat, because the percentage of body fat from raw data is more reasonable;
- **IDNO-48,76:** We delete these observations because we don’t know which part of the data is wrong;
- **IDNO-182:** We delete it because the percentage of body fat is too low;
- **IDNO-216:** We delete it because the percentage of body fat is too large.

Method 2: The formula of adiposity

BMI(Adiposity) = WEIGHT / HEIGHT^2 (kg/m^2)



IDNO	ADIPOSITY from raw data	ADIPOSITY from formula	WEIGHT	HEIGHT
42	29.9	165.6	205.00	29.50
163	24.4	27.4	184.25	68.75
221	24.5	21.7	153.25	70.50

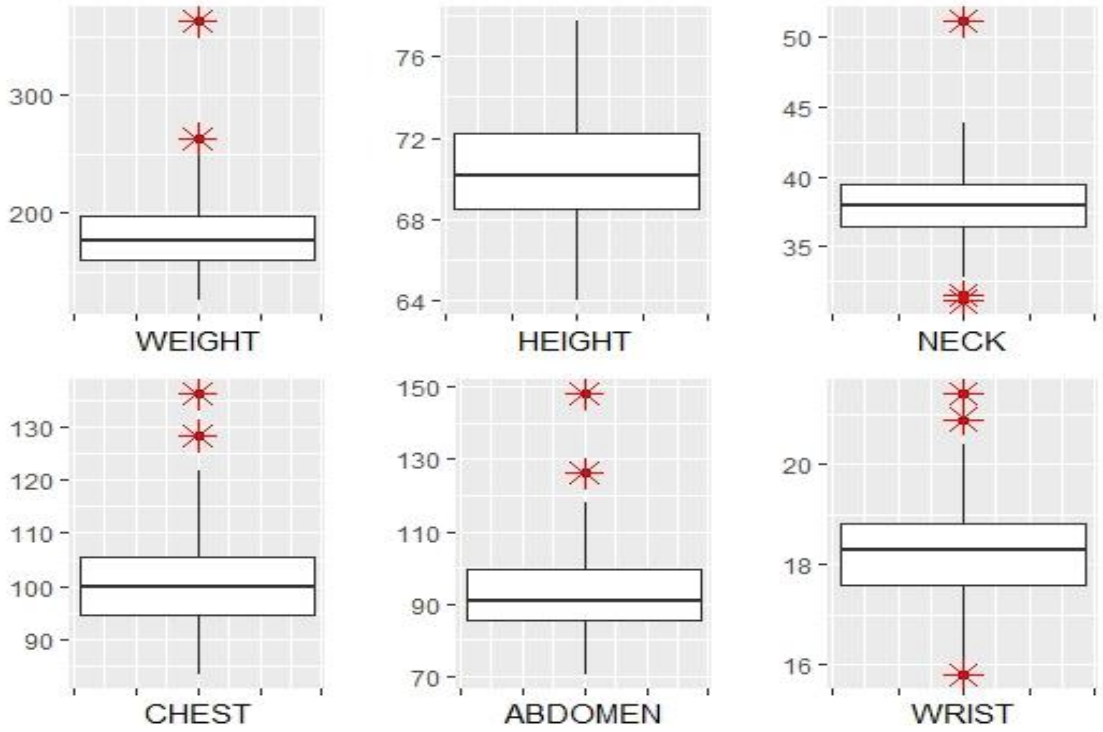
Dealing with the outliers:

We compare the predicted adiposity and the input adiposity and check whether weights and heights from raw data are correct.

- **IDNO-42:** The height in raw data is 29.50 inches, which seems incorrect. We use the BMI formula to change it as 69.45 inches (It seems that it is a writing mistake);
- **IDNO-163,221:** We delete these observations because we think they are outliers which may influence the result of following statistical model.

Step 2: Visualizing Data

Making Boxplot to find potential outliers



Potential Outliers: (IDNO-39: Extreme Case)

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
39	33.8	46	363.15	72.25	48.9	51.2	136.2	148.1	147.7	87.3	49.1	29.6	45.0	29.0	21.4
41	33.1	45	262.75	68.75	39.1	43.2	128.3	126.2	125.6	72.5	39.6	26.6	36.4	32.7	21.4
236	33.6	65	224.50	68.25	33.9	38.8	119.6	118.0	114.3	61.3	42.1	23.4	34.9	30.1	19.4
104	17.8	43	165.50	68.50	24.8	31.1	93.1	87.3	96.6	54.7	39.0	24.8	31.0	29.4	18.8
45	8.4	39	125.25	68.00	19.1	31.5	85.1	76.0	88.2	50.0	34.7	21.0	26.1	23.1	16.1
35	31.1	41	247.25	73.50	32.2	42.1	117.0	115.6	116.1	71.2	43.3	26.3	37.3	31.7	19.7
166	34.7	35	228.25	69.50	33.3	40.4	114.9	115.9	111.9	74.4	40.6	24.0	36.1	31.8	18.8
150	19.1	26	241.75	74.50	30.7	41.8	108.3	102.9	114.4	72.9	43.5	25.1	38.5	33.8	19.6
238	31.4	67	227.75	72.75	30.3	41.3	115.8	113.4	109.8	65.6	46.0	25.4	35.3	29.8	19.5
188	36.5	42	244.25	76.00	29.8	41.8	115.2	113.7	112.4	68.5	45.0	25.5	37.1	31.2	19.9
31	12.3	32	182.00	73.75	23.6	38.7	100.5	88.7	99.8	57.5	38.7	33.9	32.5	27.7	18.4
84	25.8	67	167.00	67.50	26.0	36.5	98.9	89.7	96.2	54.7	37.8	33.7	32.4	27.7	18.2
157	12.8	30	136.50	68.75	20.3	35.9	88.7	76.6	89.8	50.1	34.8	21.8	27.0	34.9	16.9
172	24.6	36	226.75	71.75	31.0	41.5	115.3	108.8	114.4	69.2	42.4	24.0	35.4	21.0	20.1
220	12.8	55	126.50	66.75	20.0	33.4	88.8	78.2	87.5	50.8	33.0	19.7	25.3	22.0	15.8

Step 3: Statistical Modeling

Model 1

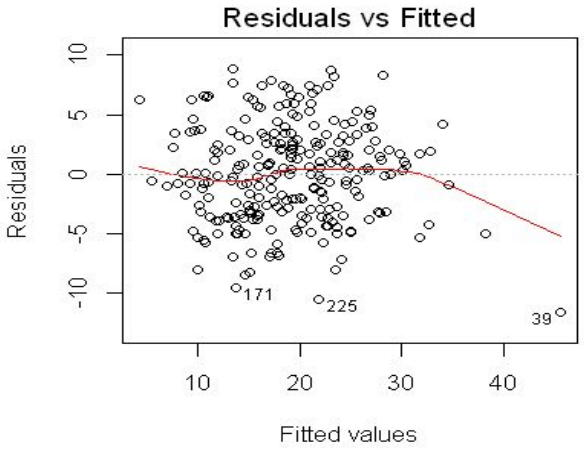
The formula below is how US army compute bodyfat for men:

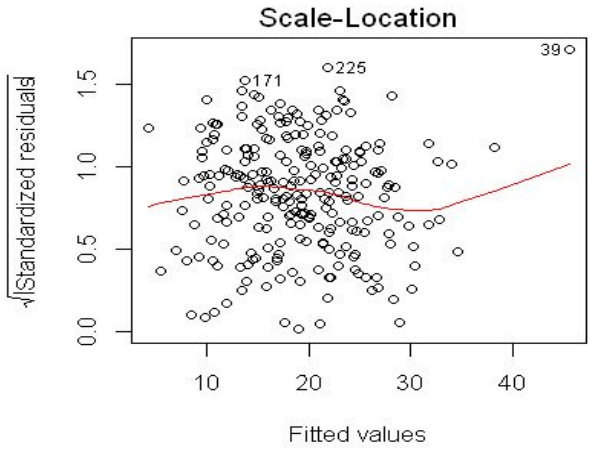
$$BodyFat \sim \log(Waist - Neck) + \log(Height)$$

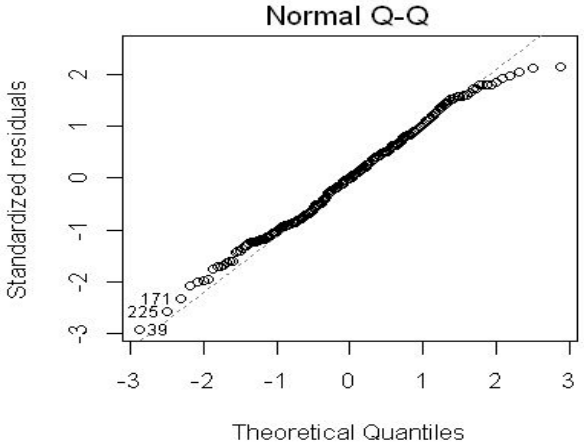
Given that data of waist are not provided and to simplify the model, Waist - Neck is replaced by Abdomen

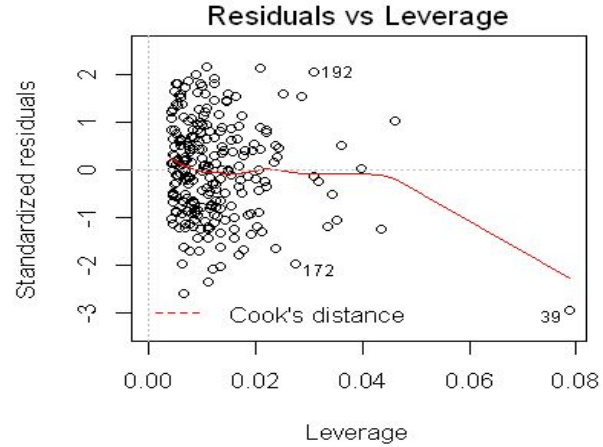
The final model is:

$$BodyFat \sim \log(Abdomen) + \log(Height)$$


Residuals vs Fitted
Residuals
Fitted values
171
225
39


Scale-Location
√|Standardized residuals|
Fitted values
171
225
39


Normal Q-Q
Standardized residuals
Theoretical Quantiles
171
225
39


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
192
172
39

Model 2

It's known that there's a strong linear relationship between body fat and reciprocal of body density. Thus, It's a good idea to estimate reciprocal of body density.
From the definition of density, it can be derived that:

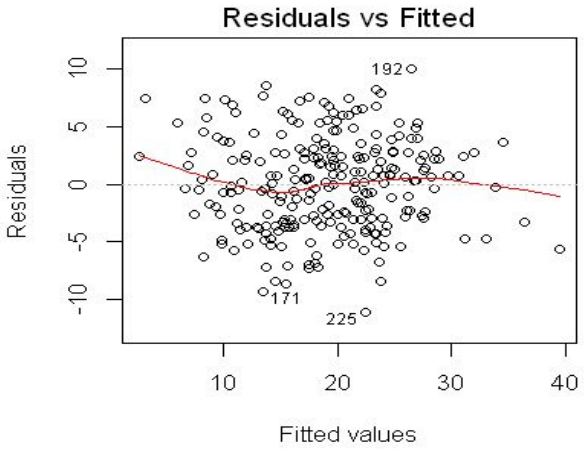
$$\rho = \frac{Weight}{Volume} = \frac{Weight}{Height \times Area} = \frac{Weight}{(Height)^2} \times \frac{Height}{\pi R^2} = BMI \times \frac{Height}{\pi \times (\frac{Abdomen}{4\pi})^2}$$

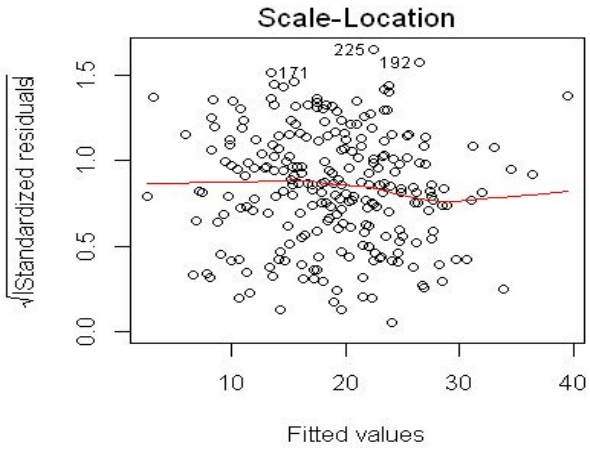
Volume of each person must equals to a certain cylinder with equal height and a certain radius r. Both radius r and BMI can be considered a linear function of abdomen. To simplify our model, we replace r with abdomen. In addition, some terms are ignored to further simplify the model, the final model is:

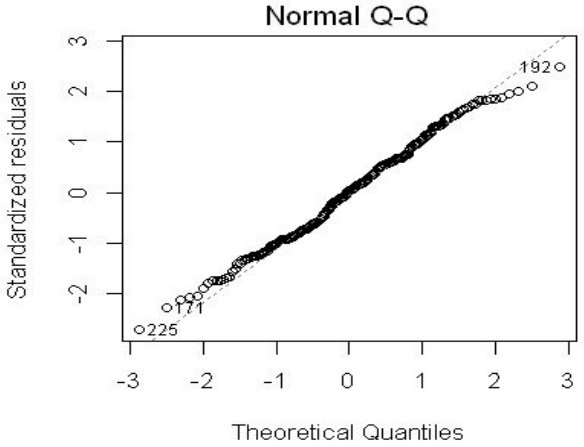
$$\rho = (k \times Abdomen + b)\frac{Height}{\frac{(Abdomen)^2}{16\pi}}$$

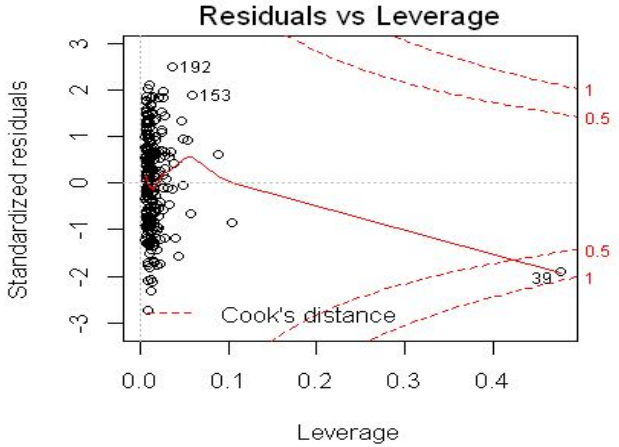
$$\frac{1}{\rho} = \frac{\frac{1}{16\pi} \times (Abdomen)^2}{(k \times Abdomen + b)Height} = \frac{1}{16\pi k^2}[\frac{(k \times Abdomen + b)}{Height} + \frac{b^2}{(k \times Abdomen + b)Height} - \frac{2b}{Height}]$$

$$Body\ fat \sim \frac{1}{\rho} \sim \frac{(k \times Abdomen + b)}{Height} + \frac{b^2}{(k \times Abdomen + b)Height} - \frac{2b}{Height}$$


Residuals vs Fitted
Residuals
Fitted values
192
171
225
10
5
0
-5
-10
10
20
30
40


Scale-Location
√|Standardized residuals|
Fitted values
225
192
171
1.5
1.0
0.5
0.0
10
20
30
40


Normal Q-Q
Standardized residuals
Theoretical Quantiles
192
171
225
3
2
1
0
-1
-2
-3


Residuals vs Leverage
Standardized residuals
Leverage
192
153
39
Cook's distance
1
0.5
0.5
1
0.0
0.1
0.2
0.3
0.4

Model 3

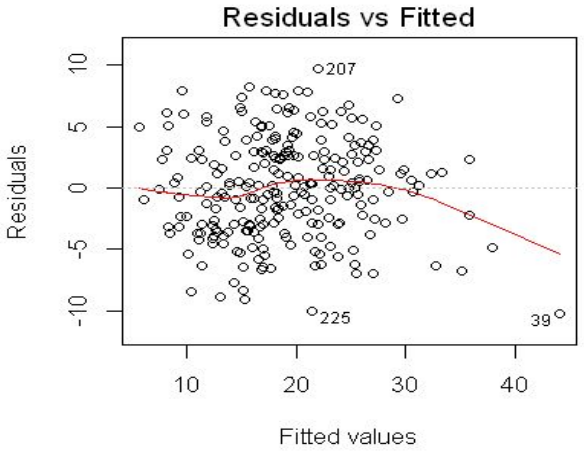
Exhaustive method is the easiest but most time-consuming way to built a good model. The result of exhaustive model is shown below.

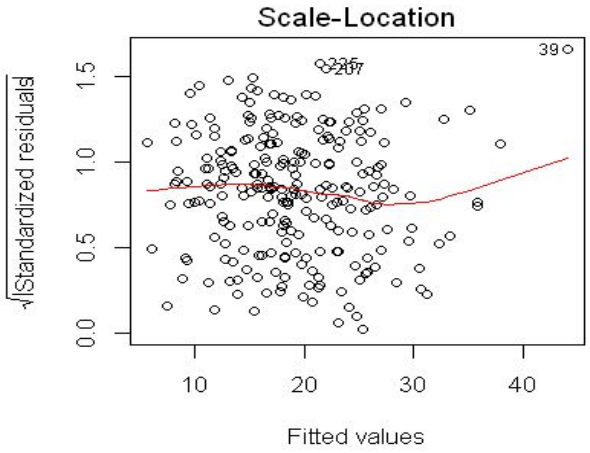
	X	mindex	n	predictors	rsquare	adjr	predrsq	cp	aic	sbic	sbc	msep	fpe	apc	hsp
1	7	1	1	ABDOMEN	6.478017e-01	0.646358275	0.63726116	73.97764	1438.179	739.0771	1448.695	20.09186	20.09053	0.3579720	0.08201031
10	13	10	1	FOREARM	1.292661e-01	0.125697564	0.11276513	539.18615	1660.844	959.0469	1671.360	49.67276	49.66947	0.8850082	0.20275270
11	14	11	1	WRIST	1.234173e-01	0.119824756	0.11011352	544.43348	1662.491	960.6796	1673.007	50.00642	50.00310	0.8909529	0.20411462
12	1	12	1	AGE	8.436381e-02	0.080611198	0.07030696	579.47066	1673.213	971.3119	1683.729	52.23431	52.23084	0.9306466	0.21320833
13	11	13	1	ANKLE	6.595369e-02	0.062125634	0.04114211	595.98744	1678.110	976.1687	1688.626	53.28455	53.28101	0.9493585	0.21749517
14	3	14	1	HEIGHT	2.208779e-07	-0.004098139	-0.01819175	655.15815	1694.895	992.8202	1705.411	57.04700	57.04321	1.0163932	0.23285262
15	32	15	2	WEIGHT ABDOMEN	7.080663e-01	0.705663519	0.69825412	21.91079	1394.013	695.5229	1408.034	16.79158	16.78880	0.2991420	0.06853935
16	84	16	2	ABDOMEN WRIST	6.868582e-01	0.684280881	0.67479081	40.93780	1411.264	712.3695	1425.286	18.01144	18.00845	0.3208737	0.07351852
17	62	17	2	NECK ABDOMEN	6.825572e-01	0.679944471	0.67141610	44.79649	1414.620	715.6480	1428.642	18.25883	18.25580	0.3252809	0.07452830
18	78	18	2	ABDOMEN HIP	6.805648e-01	0.677935729	0.66886836	46.58393	1416.159	717.1518	1430.181	18.37342	18.37037	0.3273225	0.07499605
19	43	19	2	HEIGHT ABDOMEN	6.795636e-01	0.676926292	0.66585525	47.48217	1416.929	717.9040	1430.950	18.43101	18.42795	0.3283484	0.07523111
20	80	20	2	ABDOMEN KNEE	6.649717e-01	0.662214258	0.65198367	60.57344	1427.884	728.6108	1441.905	19.27032	19.26712	0.3433006	0.07865696

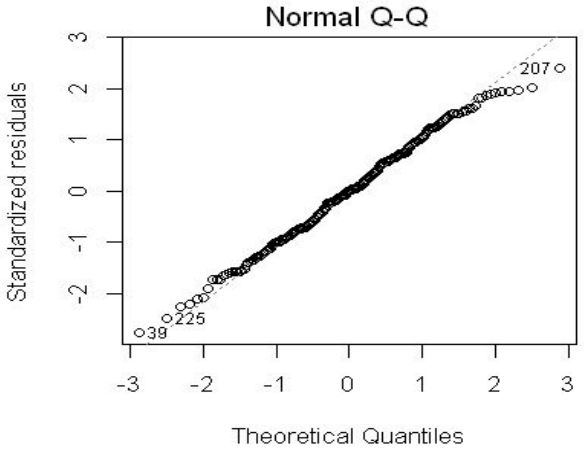
The combination with highest rsquare is Weight and abdomen.

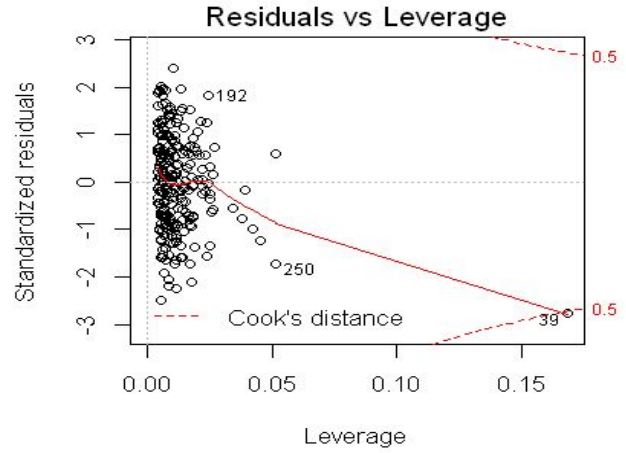
The simple model from exhaustive method is :

BodyFat~ **Abdomen + Weight**


Residuals vs Fitted
Residuals
Fitted values
207
225
39


Scale-Location
√|Standardized residuals|
Fitted values
39


Normal Q-Q
Standardized residuals
Theoretical Quantiles
207
225
39


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5
192
250
39

Model Performance

The adjusted R square of the three models are listed below:

	Model 1	Model 2	Model 3
Adjusted R square	0.6971	0.703	0.7057
Residual standard error	4.132	4.091	4.073
F - statistics	282.9	194.3	294.7

It’s hard to tell which model is better from this table.

Step 4: Model Selection and Diagnose

Model Selection

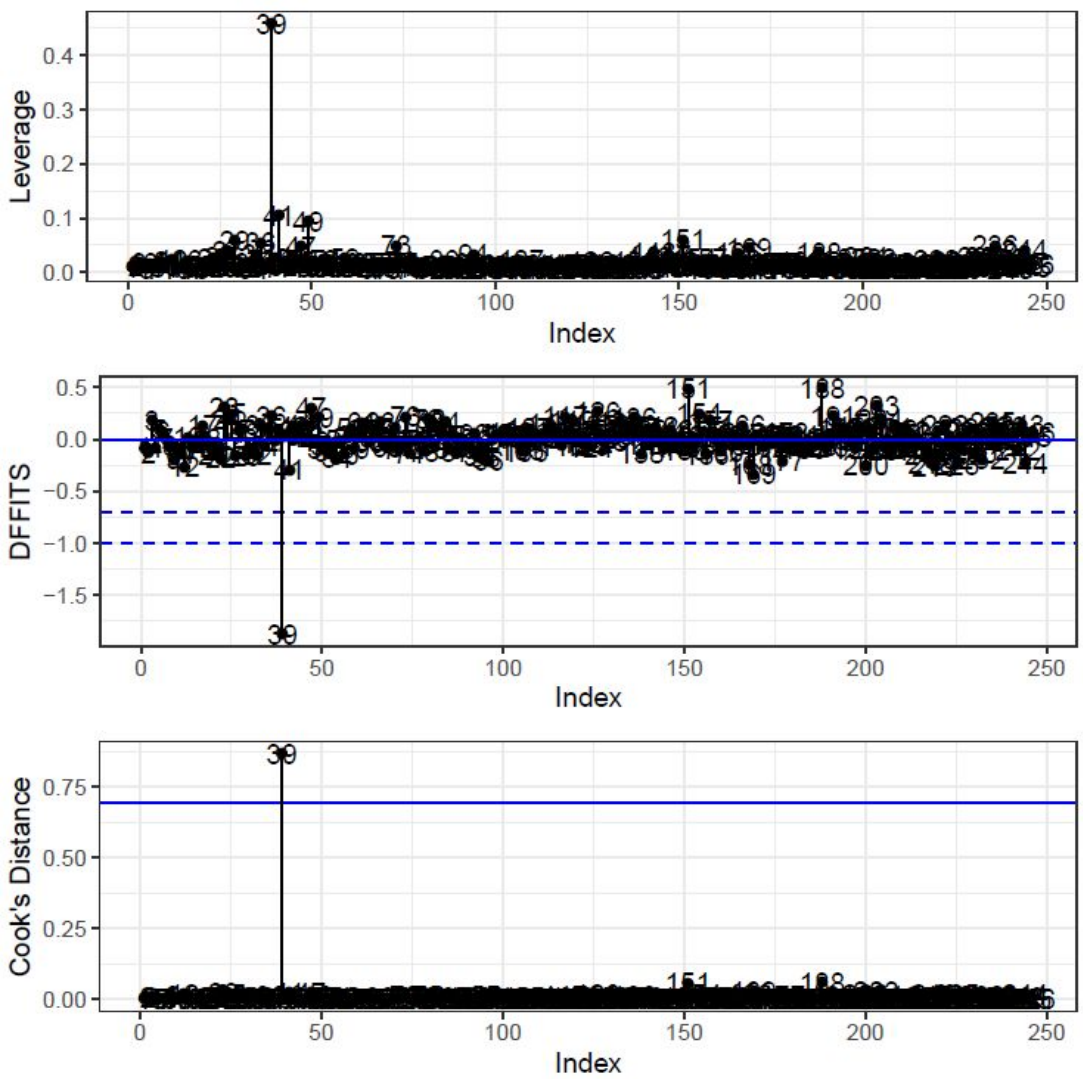
Cross validation methods are taken to do model selection.
The results are listed below:

	Model 1	Model 2	Model 3
mean of MSE	4.072	4.044	4.111
standard error of MSE	0.510	0.406	0.392

The three models are almost equally good, but model 2 and model 3 is a bit better than model 1 due to low standard error of MSE.

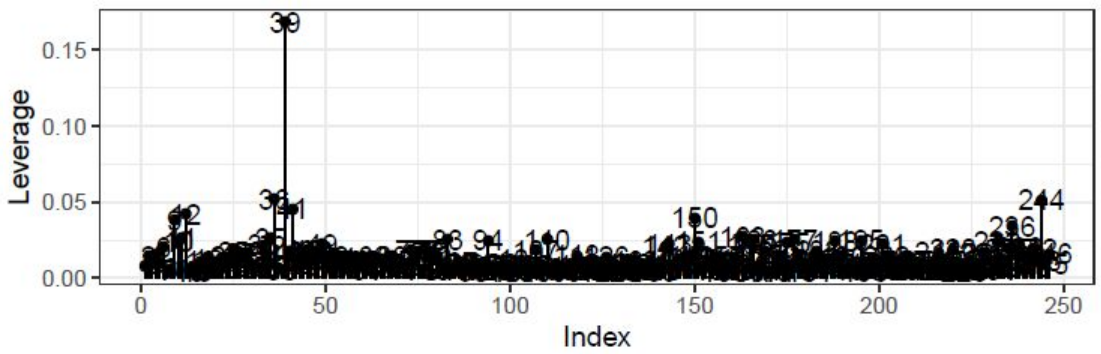
Model Diagnose

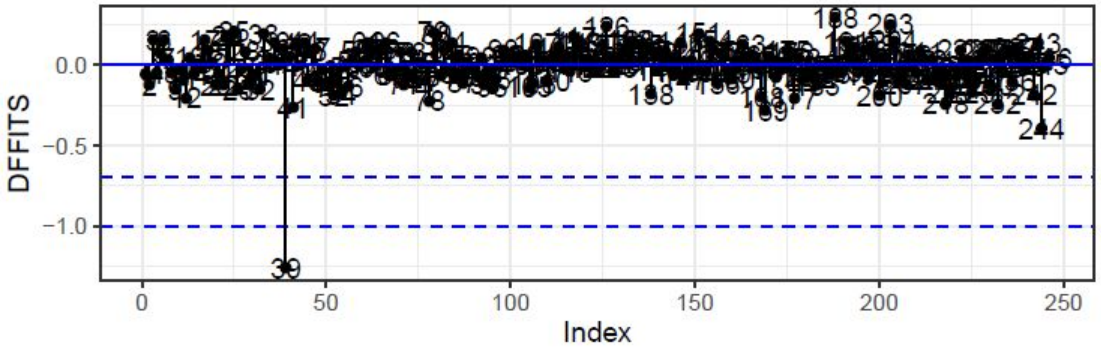
Model 2

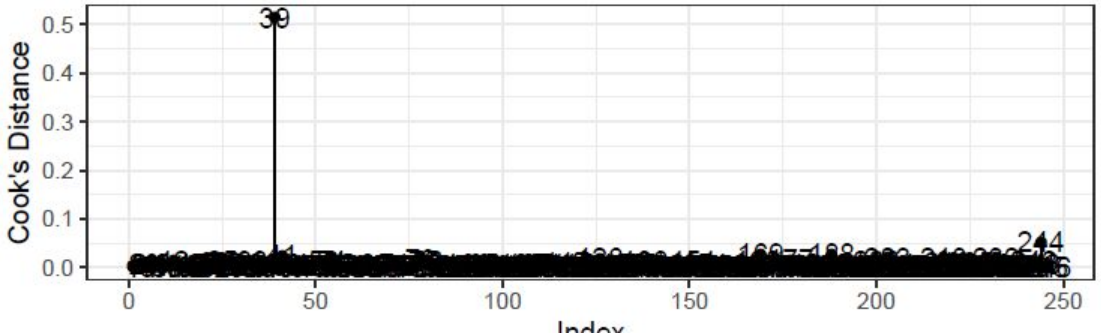


Model Diagnose

Model 3







Additional Deletetion of Data

For each of the three models, IDNO-39 seems to be a leverage point and if you take a glance on data of 39, you would also consider 39 as an outlier.

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
39	33.8	46	363.15	72.25	48.9	51.2	136.2	148.1	147.7	87.3	49.1	29.6	45	29	21.4

Because this is an extreme case with max WEIGHT, max NECK, max ADIPOSITY, max CHEST, ...,max HIP, max THIGH,..., we delete it and fit our final models: **Model 2** and **Model 3**.

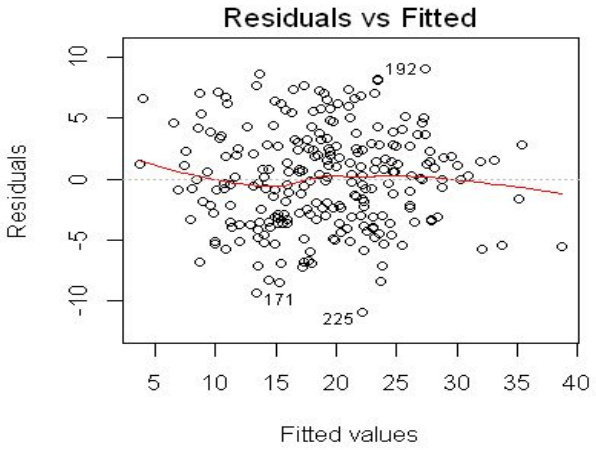
Model 2

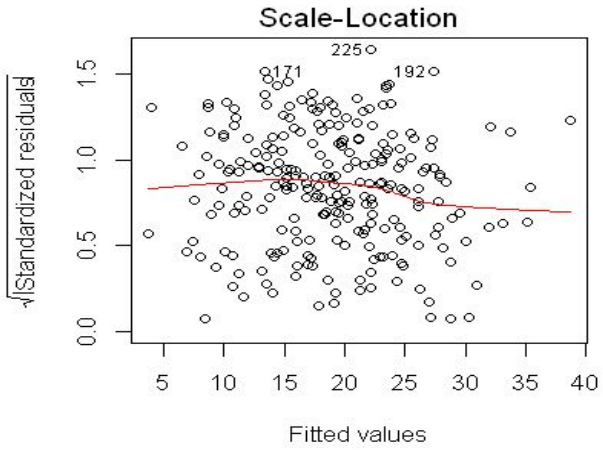
The final model 2 is:

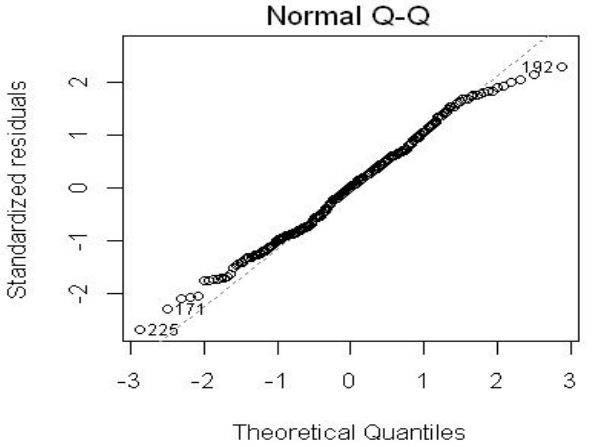
$$BodyFat = -21.62 - \frac{248324.17}{\text{Abdomen} \times \text{Height}} + 15.97\frac{\text{Abdomen}}{\text{Height}} + \frac{4095.9}{\text{Height}}$$

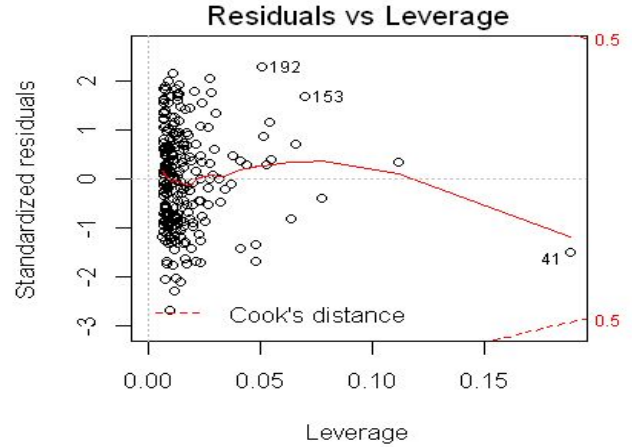
Rule of thumb:

$$\text{Your BodyFat } (\%) = -22 - \frac{24000}{\text{Your Abdomen circumference (cm)} \times \text{Your Height (inches)}} + 16\frac{\text{Your Abdomen circumference (cm)}}{\text{Your Height (inches)}} + \frac{4100}{\text{Your Height (inches)}}$$


Residuals vs Fitted
Residuals
Fitted values
192
171
225


Scale-Location
√|Standardized residuals|
Fitted values
225
171
192


Normal Q-Q
Standardized residuals
Theoretical Quantiles
192
171
225


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
192
153
41
0.5
0.5

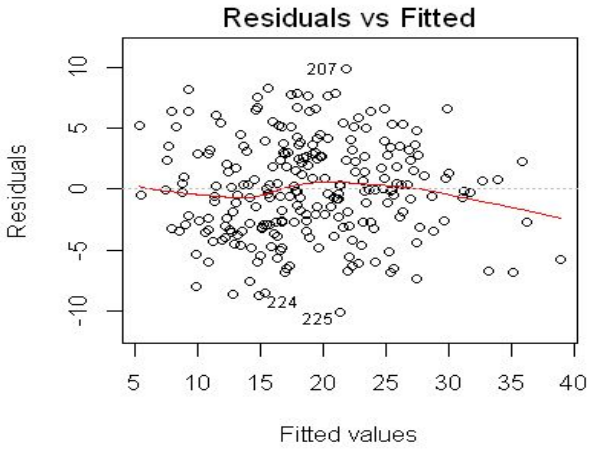
Model 3

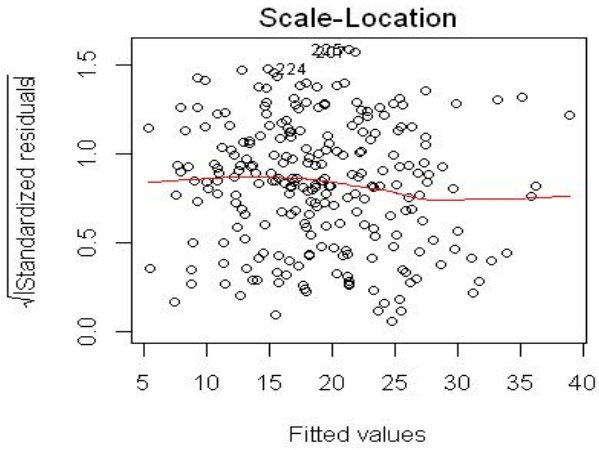
The final model 3 is:

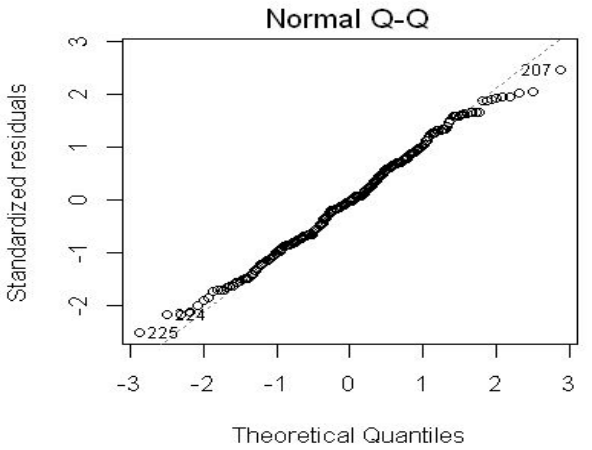
$$\textbf{Body Fat} = -41.96048 + 0.89851\textbf{Abdomen} - 0.12384\textbf{Weight}$$

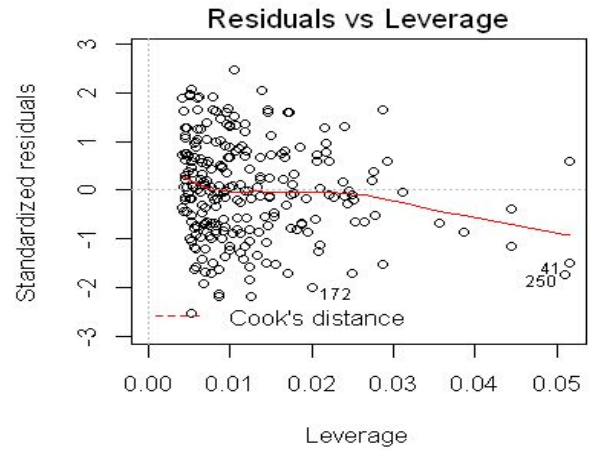
Rule of thumb:

Your Body Fat(%) = Your Abdomen circumference (cm) × 0.9
minus Your Weight (lbs) × 0.1
minus 42


Residuals vs Fitted
Residuals
Fitted values
207
224
225

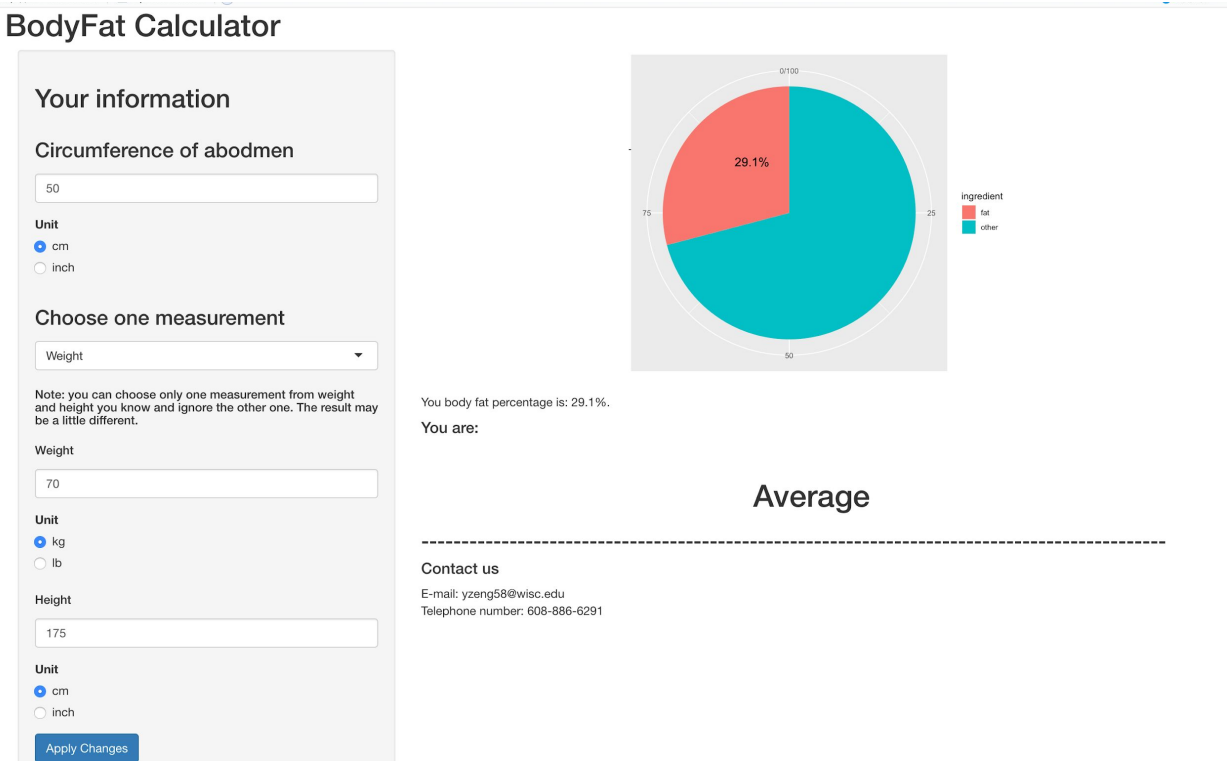

Scale-Location
√|Standardized residuals|
Fitted values
225
207
224


Normal Q-Q
Standardized residuals
Theoretical Quantiles
207
224
225


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
41
250
172

Step 5: Shiny APP

Here is the interface of our Shiny APP



Your information

Circumference of abodmen

50

Unit

- (•) cm
- ( ) inch

Choose one measurement

Weight

Note: you can choose only one measurement from weight and height you know and ignore the other one. The result may be a little different.

Weight

70

Unit

- (•) kg
- ( ) lb

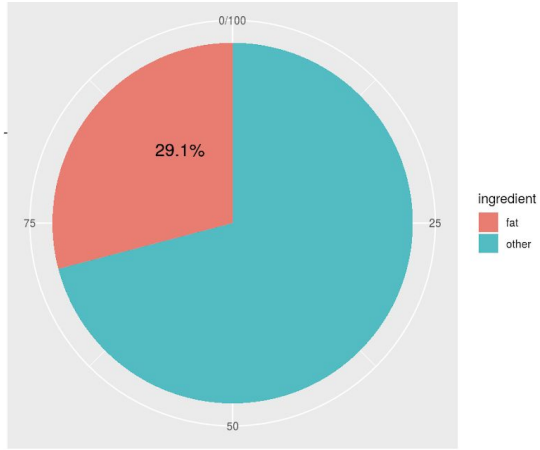
Height

175

Unit

- (•) cm
- ( ) inch

Apply Changes



You body fat percentage is: 29.1%.

You are:

Average

Contact us

E-mail: yzeng58@wisc.edu
Telephone number: 608-886-6291

Your information

Circumference of abodmen

80

Unit

- (•) cm
- ( ) inch

Choose one measurement

Height

Weight
Height

Weight

70

Unit

- (•) kg
- ( ) lb

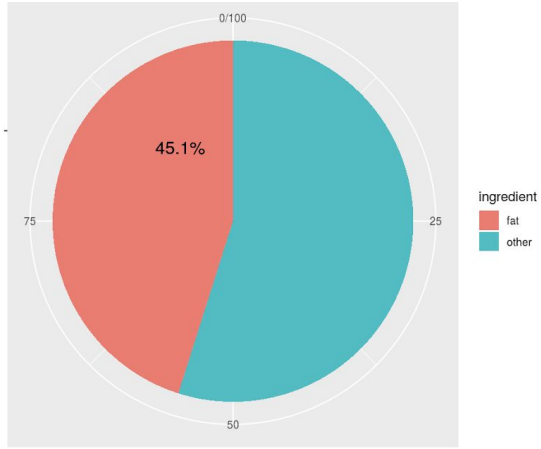
Height

175

Unit

- (•) cm
- ( ) inch

Apply Changes



You body fat percentage is: 45.1%.

You are:

Obese

Contact us

E-mail: yzeng58@wisc.edu
Telephone number: 608-886-6291

Your information

Circumference of abodmen

180

Unit

- (•) cm
- ( ) inch

Choose one measurement

Height

Note: you can choose only one measurement from weight and height you know and ignore the other one. The result may be a little different.

Weight

70

Unit

- (•) kg
- ( ) lb

Height

175

Unit

- (•) cm
- ( ) inch

Apply Changes

Your body fat is 60.9 %.

Extremely above the normal range! Please check your input.

Contact us

E-mail: yzeng58@wisc.edu
Telephone number: 608-886-6291

Strength and Weakness

Strength:

- The model we use is quite simple, which only requires two variables as input but gives a fairly good R square.
- It also provides us some flexibility that you can either input abdomen and height or abdomen and weight.

Weakness:

- The confidence interval is kind of wide.

Thank you!