How to Measure Body Fat

Liyun Zeng, Yejia Liao, Canyang Liu

STAT 628

February 4, 2019

Our Goal

Estimate percentage of body fat using clinically available measurements.

- Simple: fewer variables
- robust: Higher tolerance to perturbation
- accurate: Less MSE

Model: Linear Regression Model

What to do first?

Clean the DATA!

Let's firstly glance at the variables:

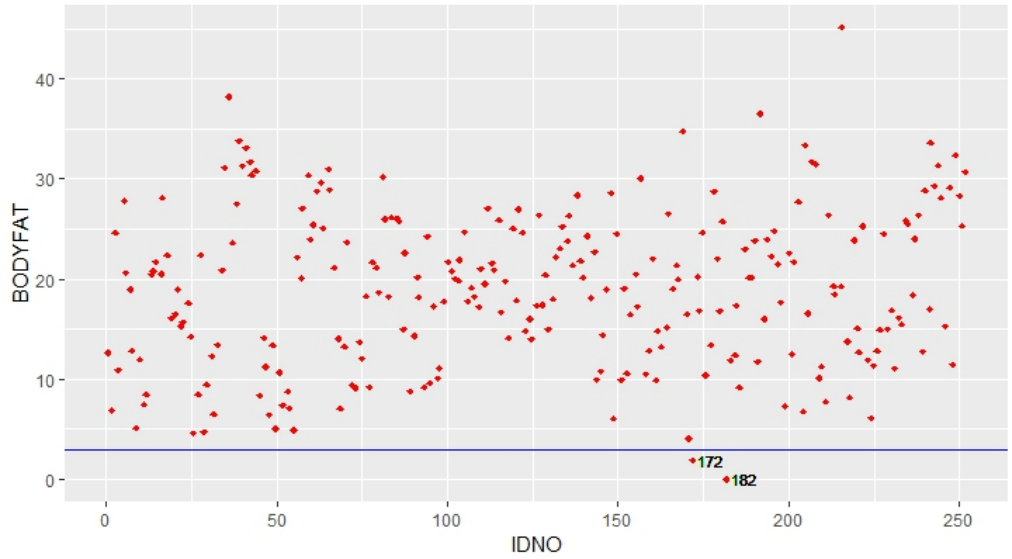
Yields	Info	Circumference
Bodyfat	Age	Neck
Density	Weight	Chest
	Height	Abdomen
	Adiposity	....

- duplicated variables that is highly correlated with other variables.
- delete them!

What to do first?

Clean the DATA!
Then, we should loot at the detailed data to find what's weird.

- Yields



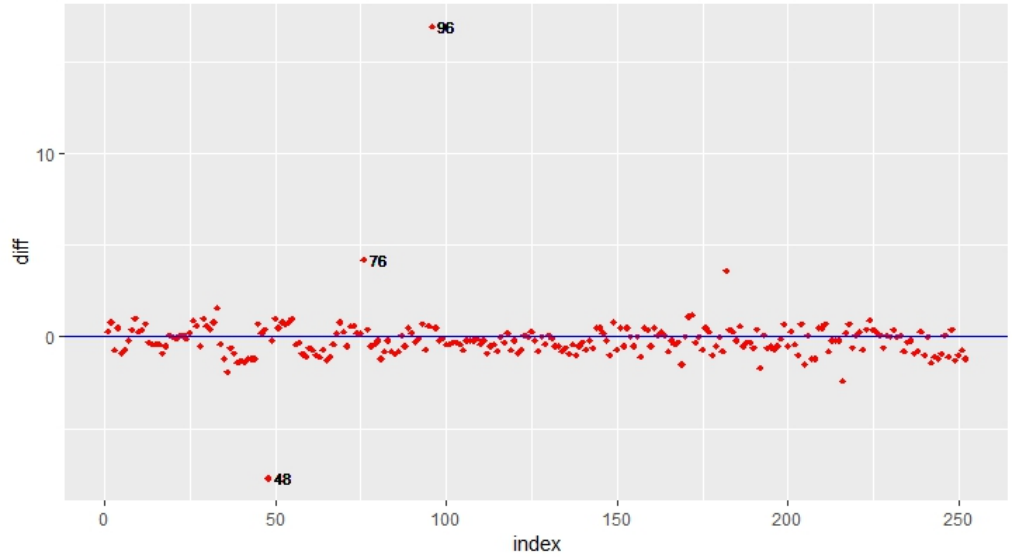
#182 and #172 are obviously wrong data, since human beings couldn't have bodyfat lower than 3% (blue line).

What to do first?

Clean the DATA!

Then, we should loot at the detailed data to find what's weird.

- Yields



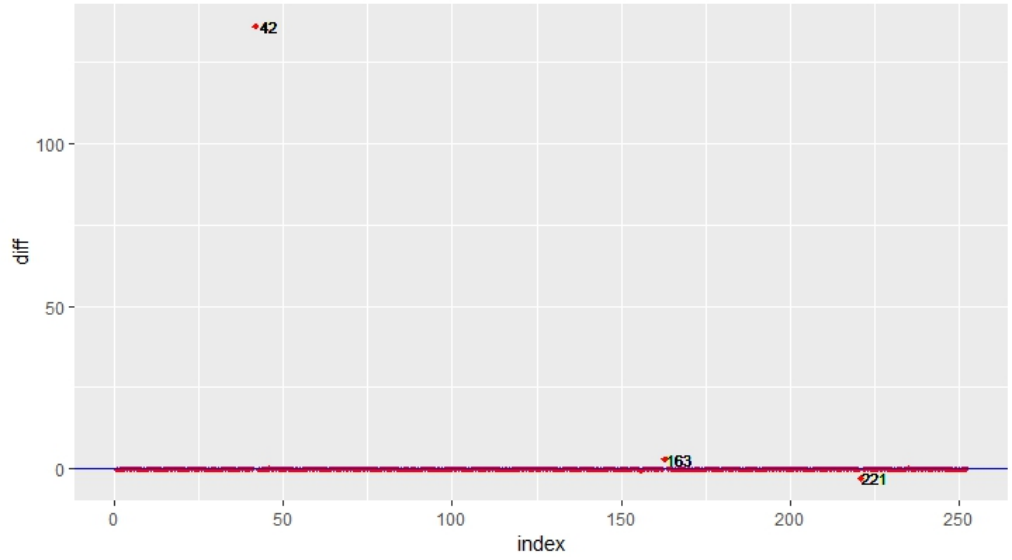
- Here we compared the estimated bodyfat using Siri's Equation and the real bodyfat
  - #96:estimated bodyfat 0.4% v.s. real bodyfat 17.3%
  - #48:estimated bodyfat 14.1% v.s. real bodyfat 6.4%
  - #76:estimated bodyfat 14.1% v.s. real bodyfat 18.3%

What to do first?

Clean the DATA!

Then, we should loot at the detailed data to find what's weird.

- Info



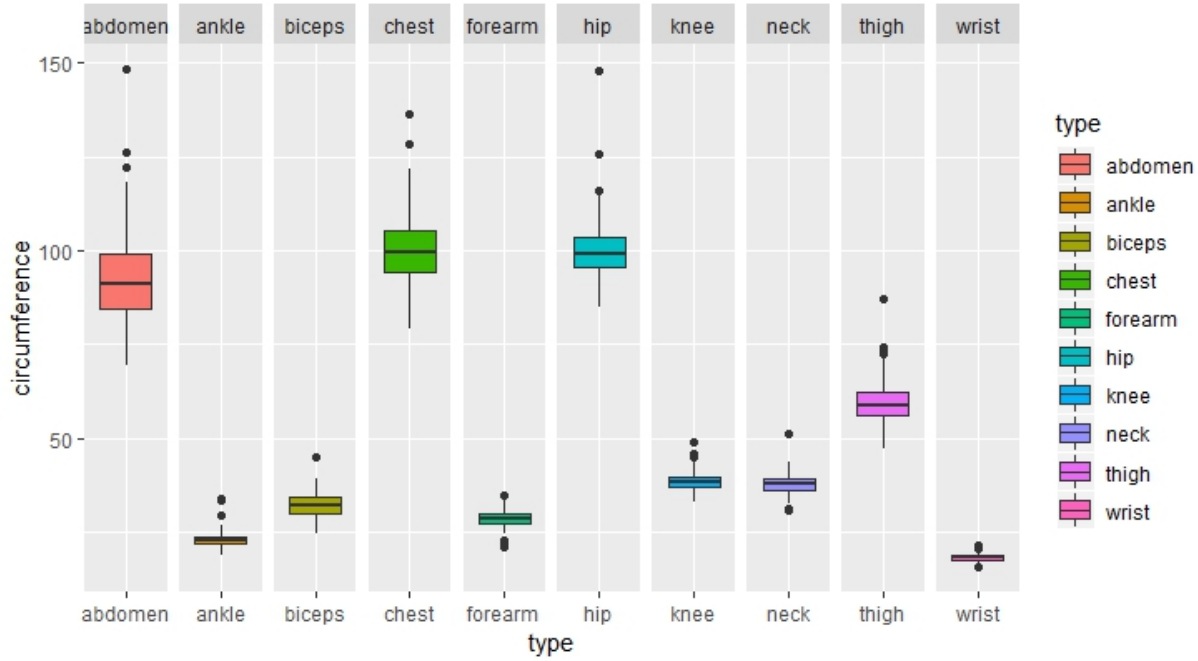
- Here we compared the estimated adiposity using bmi equation and the real adiposity
  - #42:estimated bodyfat 165.6 v.s. real bodyfat 29.9
  - #163:estimated bodyfat 27.4 v.s. real bodyfat 24.4
  - #221:estimated bodyfat 21.7 v.s. real bodyfat 24.5

What to do first?

Clean the DATA!

Then, we should look at the detailed data to find what's weird.

- Circumference



Lots of outliers, some of them are strange but reasonable, while some are weird.

What to do first?

Clean the DATA!

Then, we should look at the detailed data to find what's weird

- Circumference
  - #39 This guy is 363 pounds, correct data, but since 9 out of 10 attributes are outliers, this data has great influence to the model.
  - #41 This guy is 262 pounds, correct data, but since 4 out of 10 attributes are outliers, this data has great influence to the model.
  - #86 This guy has a normal body, but extreme big ankle. His ankles are even thicker than his knee and neck.
  - #175 This guy is 226 pounds, but his forearm is 21cm wide, even thinner than a slim lady

Data Clean Sum Up

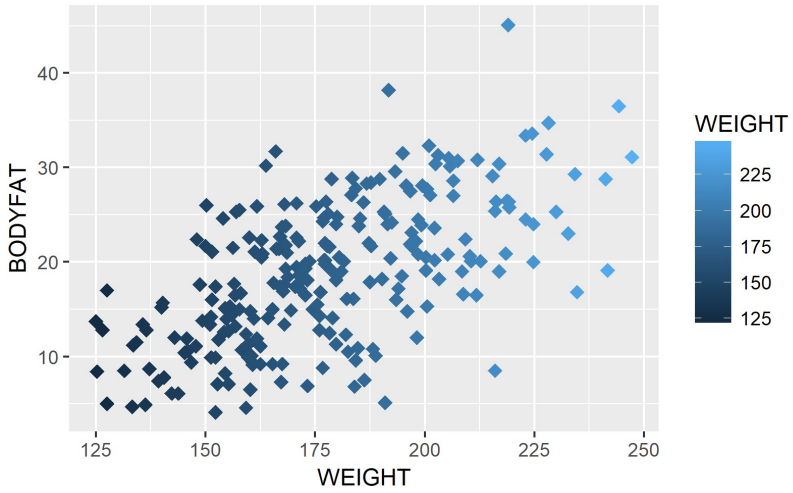
- obvious wrong data
  - #182: a skeleton (0 bodyfat)
  - #172: a skeleton (1.9 bodyfat)
  - #86: extra-big-ankle
  - #175: extra-small-forearm
- strange data
  - #96: wrong calculation of bodyfat (should be a skeleton)
  - #48: wrong calculation of bodyfat (don't know why)
  - #76: wrong calculation of bodyfat (don't know why)
  - #42: wrong calculation of adiposity (29 inches tall)
  - #163: wrong calculation of adiposity (don't know why)
  - #221: wrong calculation of adiposity (don't know why)
  - #39: extra-obese
  - #41: extra-obese
- Why don't we rewrite those data?

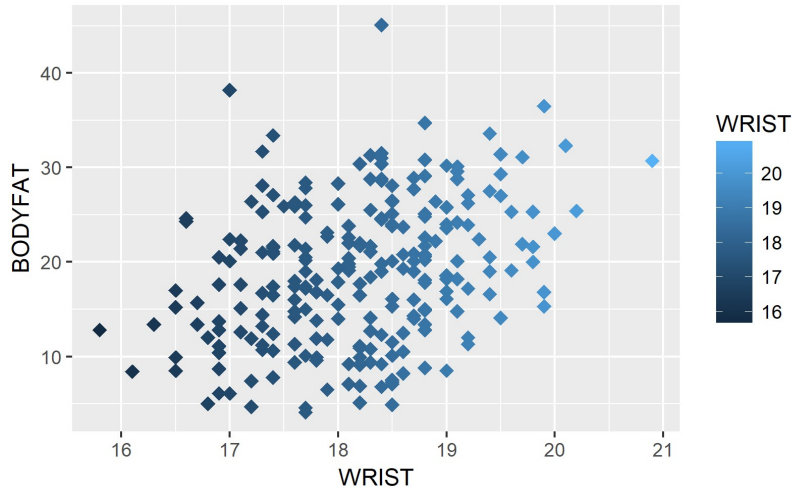
Data Clean Sum Up

- We don't know where does the wrong calculation come from.
- We want to obtain a higher accuracy for more people, so we abnegate sort of high leverage points.
- the data size is large enough to abandon 12 out of 252 data.
- We deleted all those 12 data from the dataset.

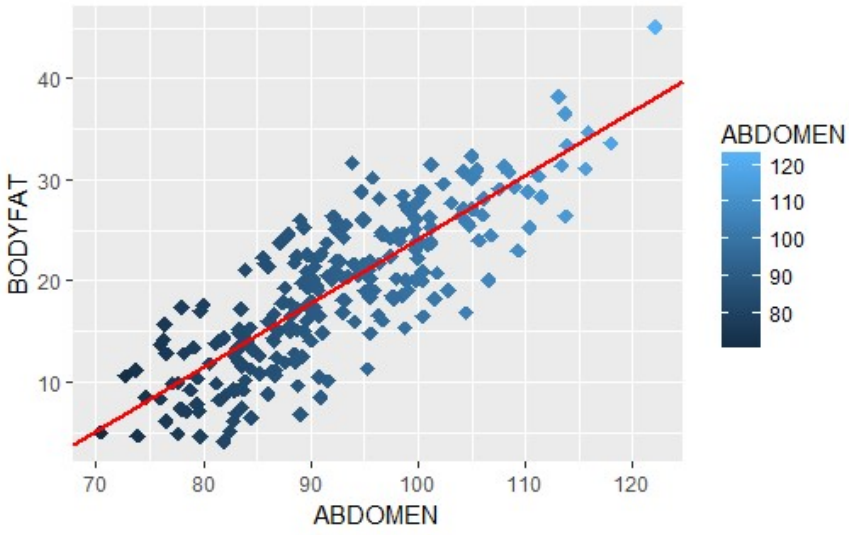
What's do next?

Let's see the scatter plot of some intuitively important variables to explore the trend using the cleaned data.



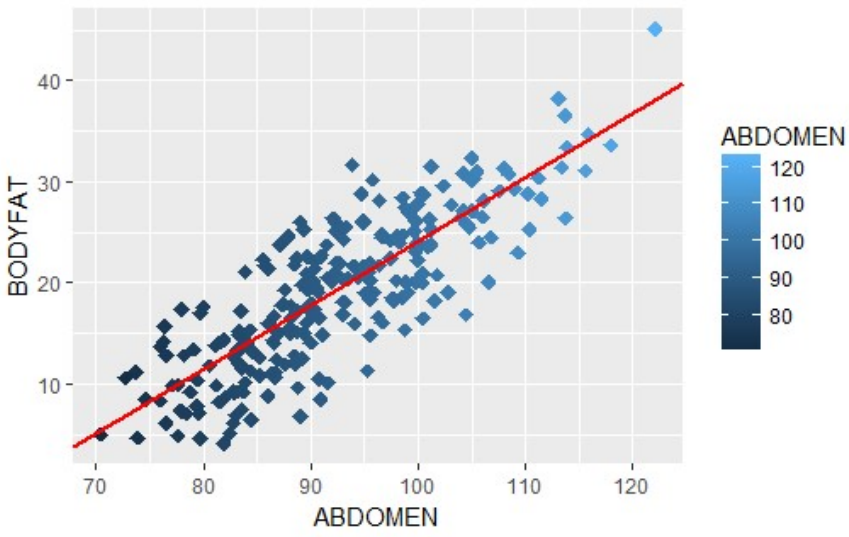


What's do next?



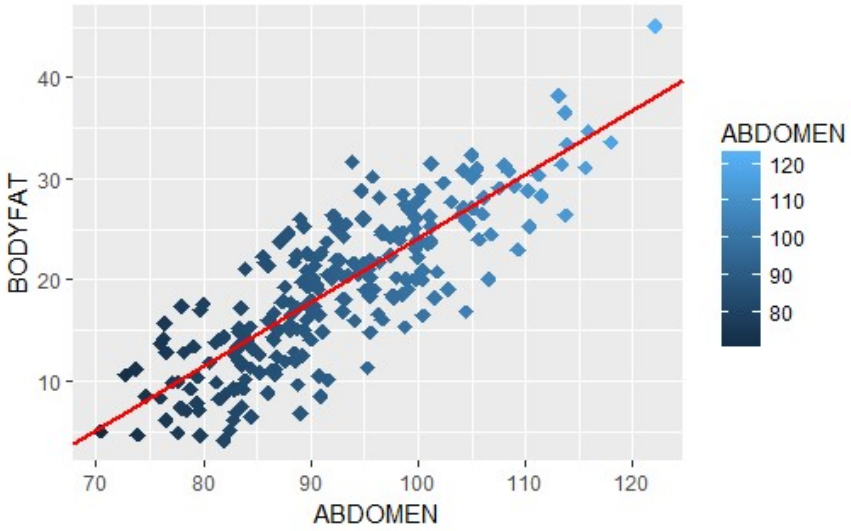
- Intuitively, we established a simple linear regression model using the circumference of Abdomen to explain the bodyfat

What's do next?



- Intuitively, we established a simple linear regression model using the circumference of Abdomen to explain the bodyfat
- Bodyfat= -38.8766 + 0.626 * Abdomen

What's do next?



- Intuitively, we established a simple linear regression model using the circumference of Abdomen to explain the bodyfat
- Bodyfat= -38.8766 + 0.626 * Abdomen

- explain 67.8% of variation in bodyfat (based on R^2)
- Mean Square Error = 18.17
- simple, but not robust, and not accurate
- Rule of Thumb: Bodyfat=0.6*Abdomen -39

Model Improvement

- Multi-variate Regression Model

	No. of variates	MSE	R^2	name of var
AIC	4	14.83	0.737	?
BIC	3	15.04	0.733	Weight, Abdomen, Wrist

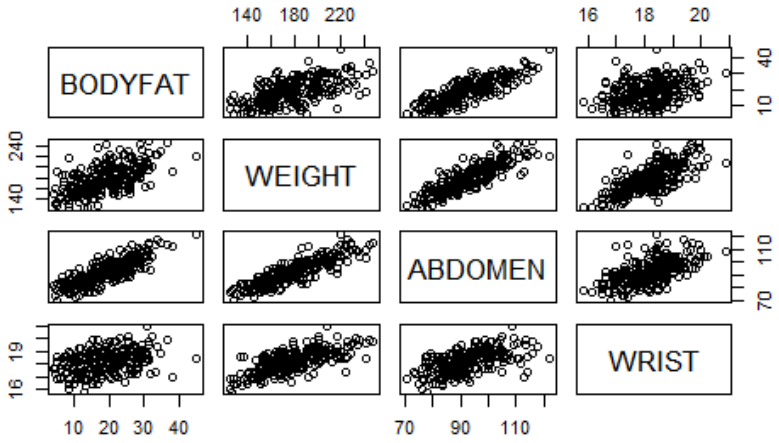
Model Improvement

- Multi-variate Regression Model

	No. of variates	MSE	R^2	name of var
AIC	4	14.83	0.737	Weight, Abdomen, Wrist, Forearm
BIC	3	15.04	0.733	Weight, Abdomen, Wrist

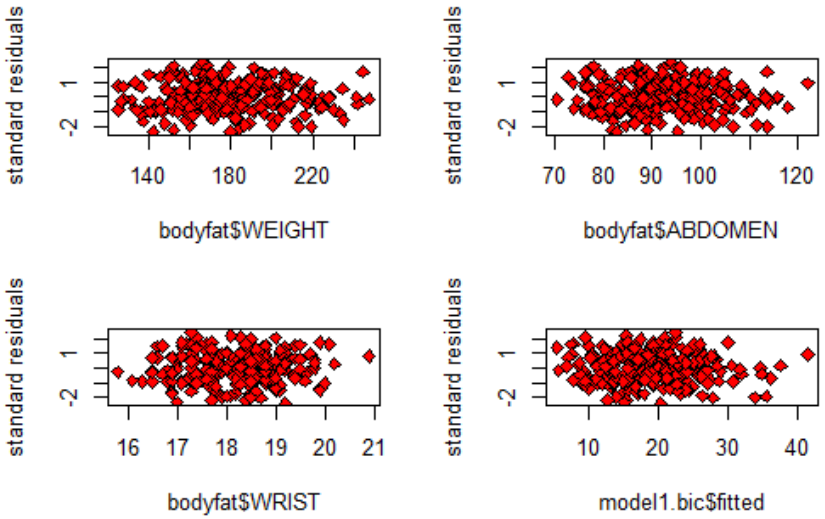
- Bic Model: Bodyfat ~ Weight+Abdomen+Wrist
- Bodyfat=0.895*Abdomen-1.109*Wrist-0.09*Weight-26.8

diagnosis



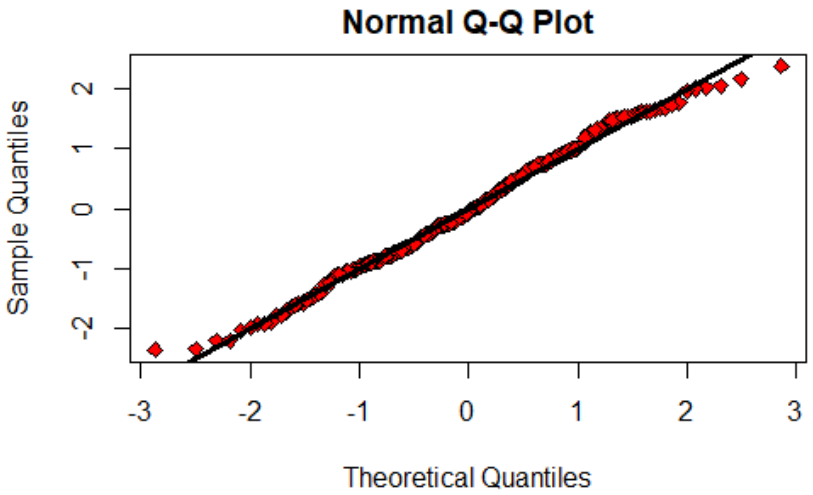
- linearity ✓
- independency ✓
- homoscedasticity ✓
- normality ✓
- multicolinearity ×

diagnosis



- linearity ✓
- independency ✓
- homoscedasticity ✓
- normality ✓
- multicolinearity ×

diagnosis



- linearity ✓
- independency ✓
- homoscedasticity ✓
- normality ✓
- multicolinearity ×

diagnosis

```
  WEIGHT  ABDOMEN    WRIST
5.331620 3.974230 2.017671

Call:
omcdiag(x = x, y = bodyfat$BODYFAT)


Overall Multicollinearity Diagnostics

                       MC Results detection
Determinant |X'X|:         0.1261         0
Farrar Chi-Square:       487.3129         1
Red Indicator:             0.7239         1
Sum of Lambda Inverse:    11.3235         0
Theil's Method:            0.5986         1
Condition Number:         79.7111         1

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

- linearity ✓
- independency ✓
- homoscedasticity ✓
- normality ✓
- multicolinearity ×

diagnosis

- robustness test
  - using different sub dataset (bootstrap sub-sample) ✓
  - using full dataset (containing those weird data) ✓

diagnosis

- robustness test
  - using different sub dataset ✓
  - using full dataset (containing those weird data) ✓
  - substitute dependent variable by "DENSITY" ✓
  - substitute explanatory variables ✓

Conclusion

- final model:
- Bodyfat=0.895*Abdomen-1.109*Wrist-0.09*weight-26.8
- Rule of Thumb
- Bodyfat=Abdomen-Wrist-0.1*weight-26.8

Conclusion

- final model:
  Bodyfat=0.895*Abdomen-1.109*Wrist-0.09*weight-26.8
- Rule of Thumb: Bodyfat=Abdomen-Wrist-0.1*weight-26.8
- simple
- robust
- accurate
- shiny `https://lzeng32.shinyapps.io/shinyapp/`

Weakness

- abnegate too many observations, may lose some important information.
- exists multi-colinearity in pairs "Weight-Abdomen", but VIF indicates that multi-colinearity is not severe.
- The accuracy still has room for improvement.
- .....

Thanks!