

How to Predict Bodyfat

Tuesday Group 2

GUANG LI, YI-HSIUAN TSAI, HAOXIANG WEI, DIAOCHU YU

INDEX

- **WHAT IS BODYFAT**
- **DATA PRE-PROCESSING**
- **MODEL BUILDING**
- **RULE OF THUMB**
- **PROS AND CONS**

WHAT IS BODYFAT

- $BODYFAT = \frac{m(FAT)}{m(BODY)} \times 100\%$
- *The body fat values we use is from Siri's (1956) equation.*

Siri's (1956) equation

Assume that the body consists of two components - lean body tissue and fat tissue.

D = Body Density (gm/cm^3)

A = proportion of lean body tissue

B = proportion of fat tissue ($A+B=1$)

a = density of lean body tissue (gm/cm^3)

b = density of fat tissue (gm/cm^3)

Solving for B we find

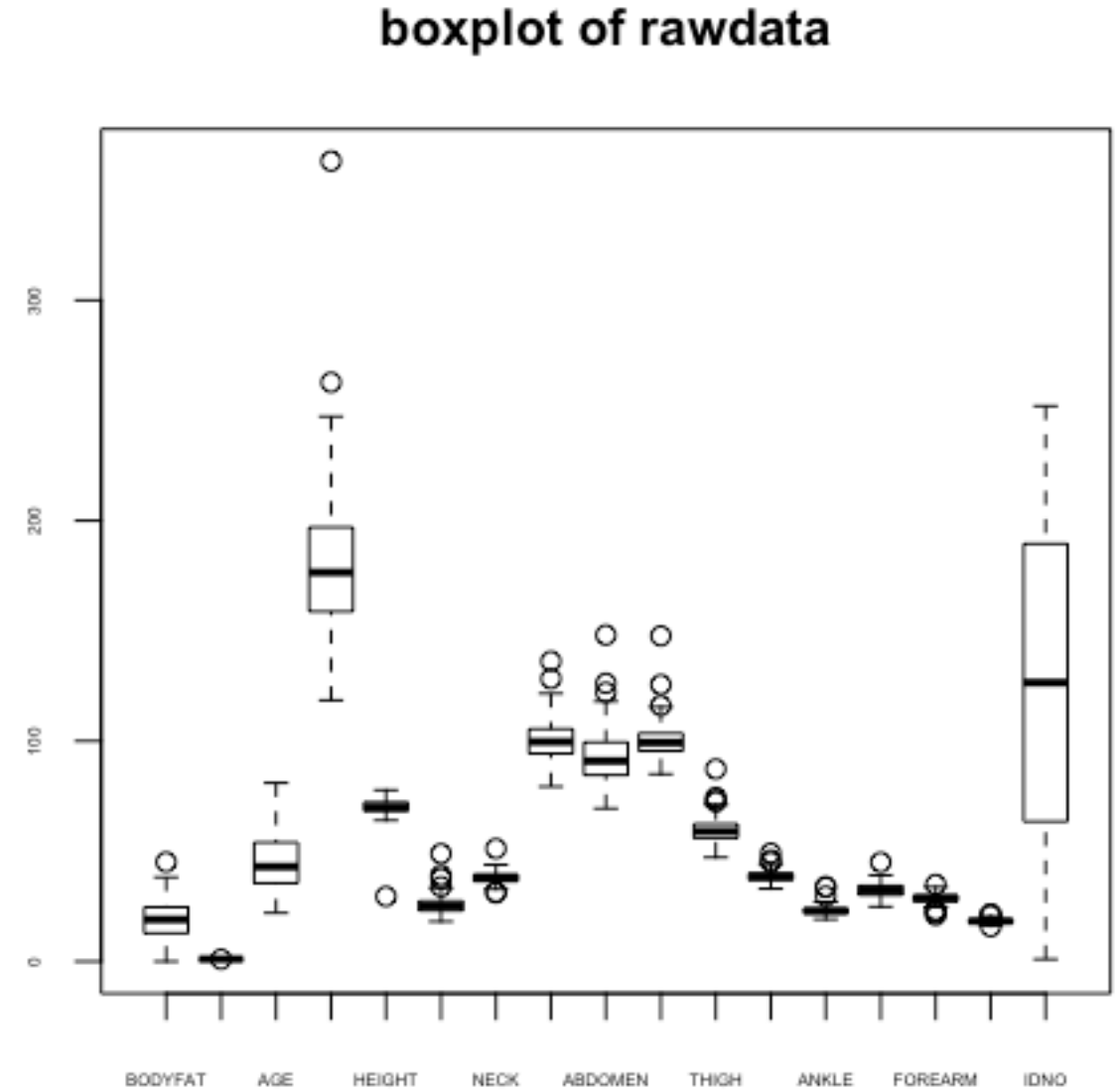
$$B = (1/D) * [ab/(a-b)] - [b/(a-b)].$$

DATA PRE- PROCESSING

- Use the linear relationship between density and bodyfat(Siri's equation), BMI formula, cook's distance to find out abnormal observations.
- Delete those observations.
- Scale the variables.

Boxplot

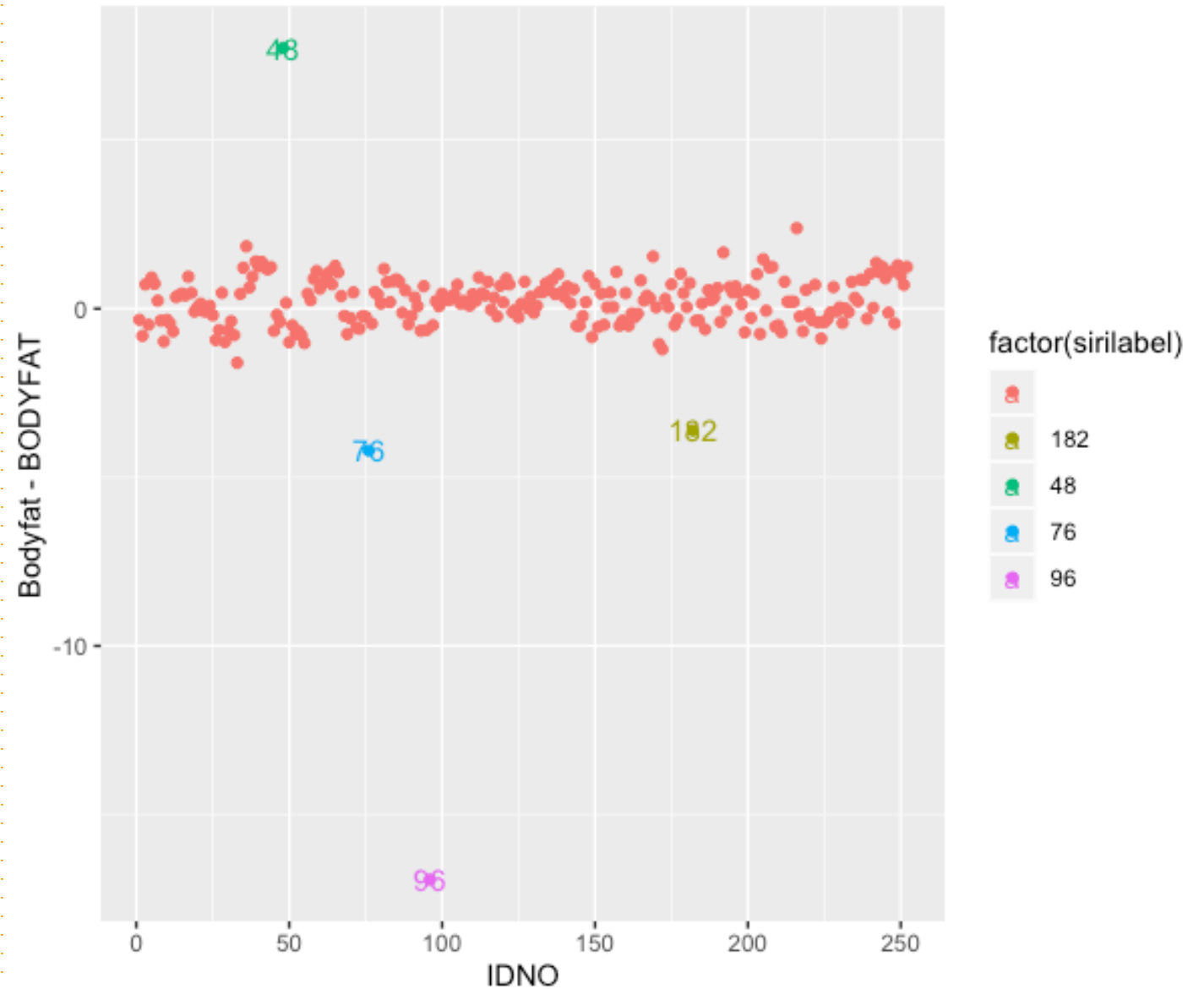
- *An obvious outlier with extremely large weight. (No. 39)*
- *Need to normalize the data, for they are of totally different scale.*



Siri's equation

$$\text{Bodyfat} = \left(\frac{495}{\text{Density}} - 450 \right) * 100\%$$

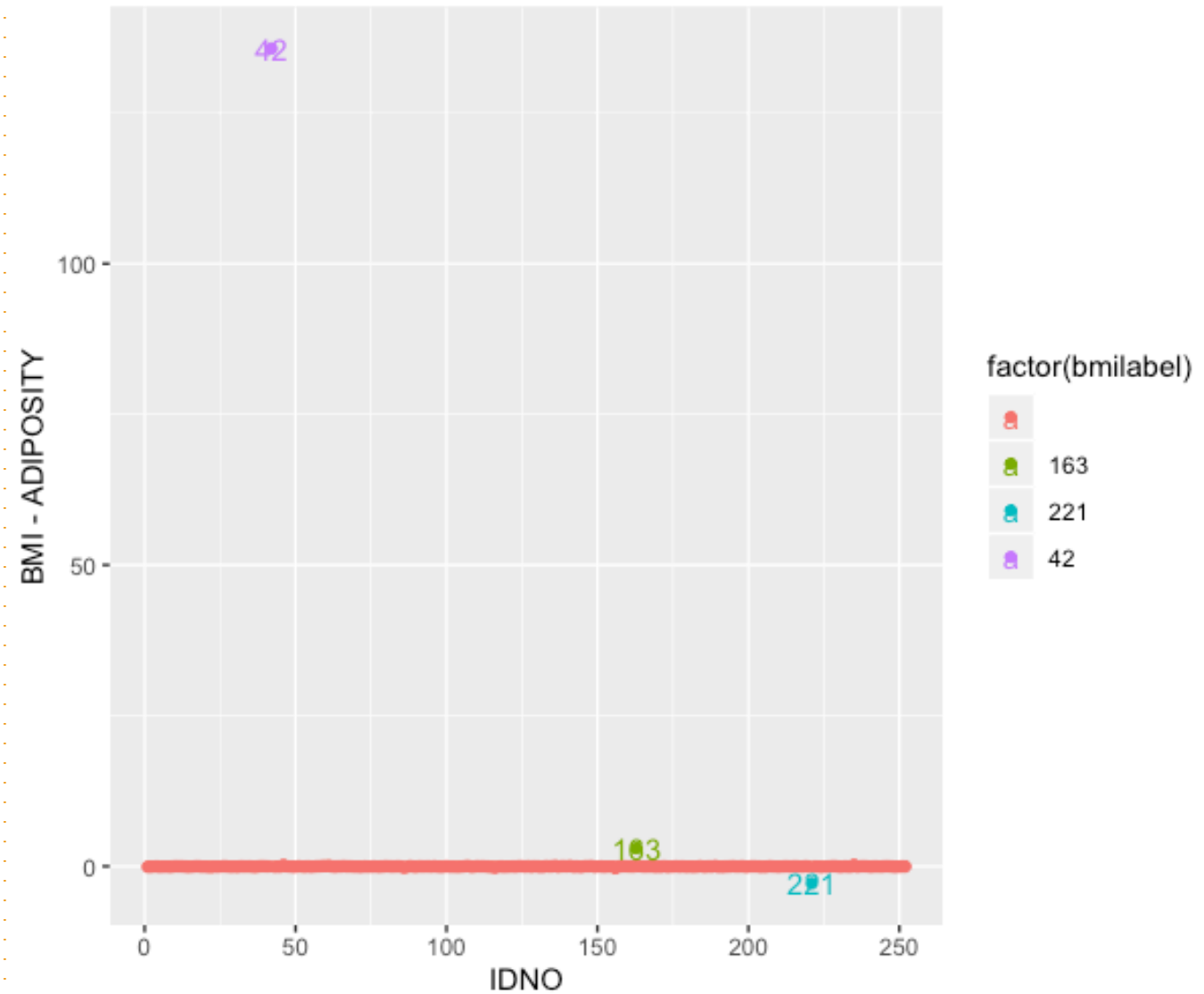
No. 48, 76, 96, 182



BMI formula

$$ADIPOSITY(BMI) = 703 \times \frac{WEIGHT(lbs)}{HEIGHT(inches)^2}$$

No. 42,163,221



Outliers

	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM
39	33.8	1.0202	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
42	31.7	1.0250	44	205.00	29.50	29.9	36.6	106.0	104.3	115.5	70.6	42.5	23.7	33.6	28.7
48	6.4	1.0665	39	148.50	71.25	20.6	34.6	89.8	79.5	92.7	52.7	37.5	21.9	28.8	26.8
76	18.3	1.0666	61	148.25	67.50	22.9	36.0	91.6	81.8	94.8	54.5	37.0	21.4	29.3	27.0
96	17.3	1.0991	53	224.50	77.75	26.1	41.1	113.2	99.2	107.5	61.7	42.3	23.2	32.9	30.8
163	13.3	1.0690	33	184.25	68.75	24.4	40.7	98.9	92.1	103.5	64.0	37.3	23.5	33.5	30.6
182	0.0	1.1089	40	118.50	68.00	18.1	33.8	79.3	69.4	85.0	47.2	33.5	20.2	27.7	24.6
221	12.7	1.0706	54	153.25	70.50	24.5	38.5	99.0	91.8	96.2	57.7	38.1	23.9	31.4	29.9

- 39 has too large size.
- Observations 48, 76, 96, 182 do not match the relationship between bodyfat and density (i.e. Siri's equation).
- Observations 42, 163, 221 do not match the BMI formula.

Model Building

- *Variable selection*
- *PCA (hard to interpret and use)*
- *CART decision tree (dataset is too small)*

Variable Selection

- Stepwise selection(AIC,BIC), Mallows cp, and lasso.
- AIC backward: ***BODYFAT ~ AGE + ADIPOSITY + NECK + ABDOMEN + HIP + THIGH+FOREARM + WRIST***
- AIC forward/both sides: ***BODYFAT ~ ABDOMEN + WEIGHT + WRIST + BICEPS***
- BIC forward/backward/both sides: ***BODYFAT ~ ABDOMEN + WEIGHT + WRIST***
- Mallows cp: ***BODYFAT ~ AGE + WEIGHT + HEIGHT + +ADIPOSITY + NECK + ABDOMEN + HIP + THIGH + BICEPS + FOREARM + WRIST***
- Adjusted R square: ***BODYFAT~AGE +WEIGHT+ HEIGHT +ADIPOSITY+NECK+CHEST+ ABDOMEN +HIP+ THIGH+BICEPS+FOREARM + WRIST***
- Lasso: ***BODYFAT ~ AGE + HEIGHT + ABDOMEN + WRIST***

WRIST or WEIGHT

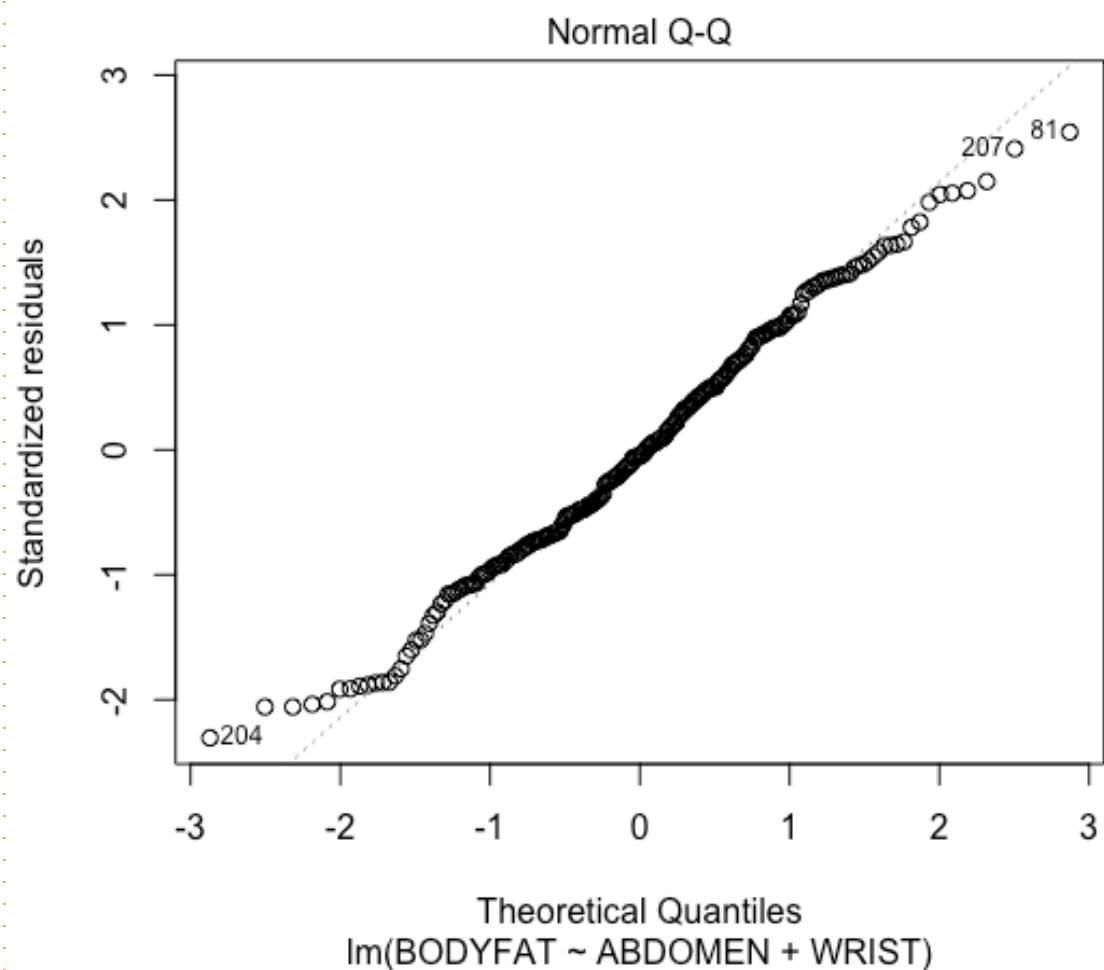
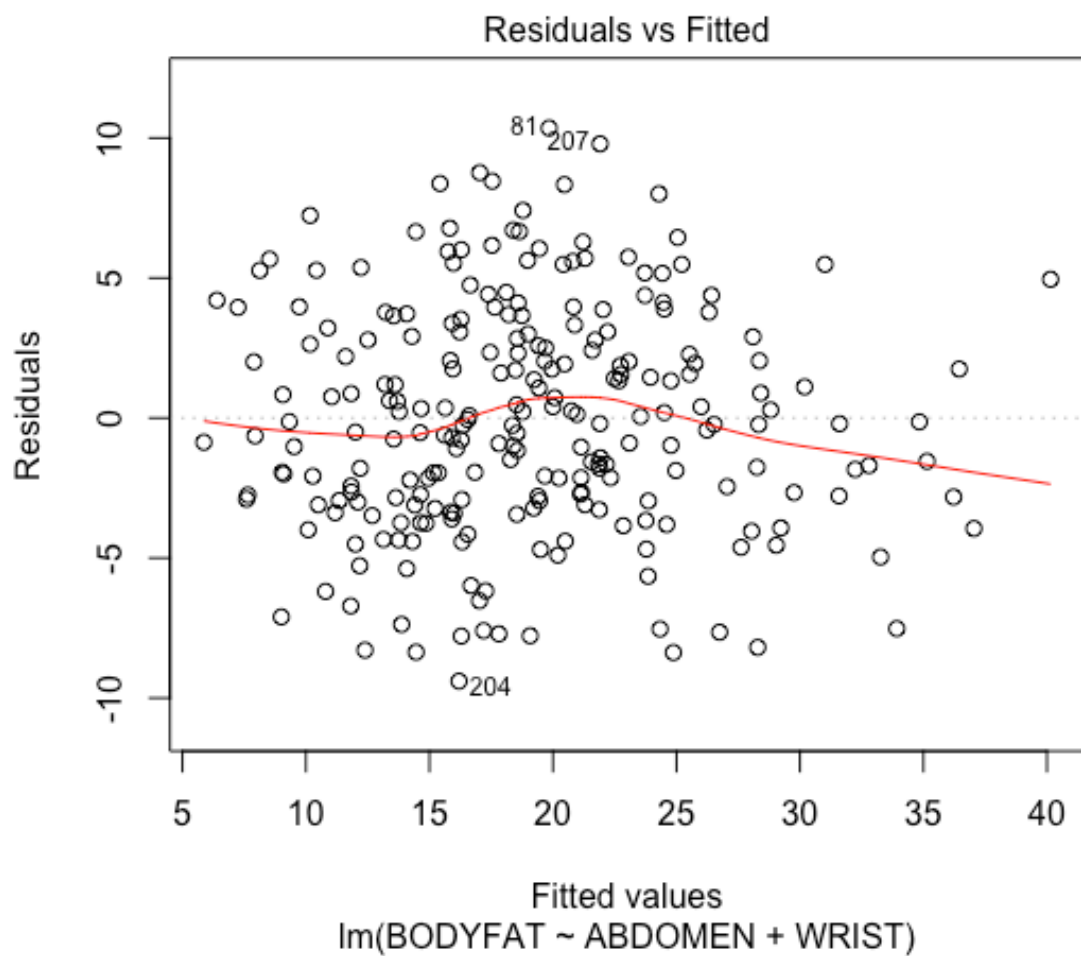
- ***ABDOMEN*** is the most important predictor beyond doubt, followed by ***WRIST*** and ***WEIGHT***.
- Model 1: ***BODYFAT ~ ABDOMEN + WEIGHT***
- Model 2: ***BODYFAT ~ ABDOMEN + WRIST***

	Model 1	Model 2
Correlation coefficient	0.603	0.873
VIF test	4.194	1.572

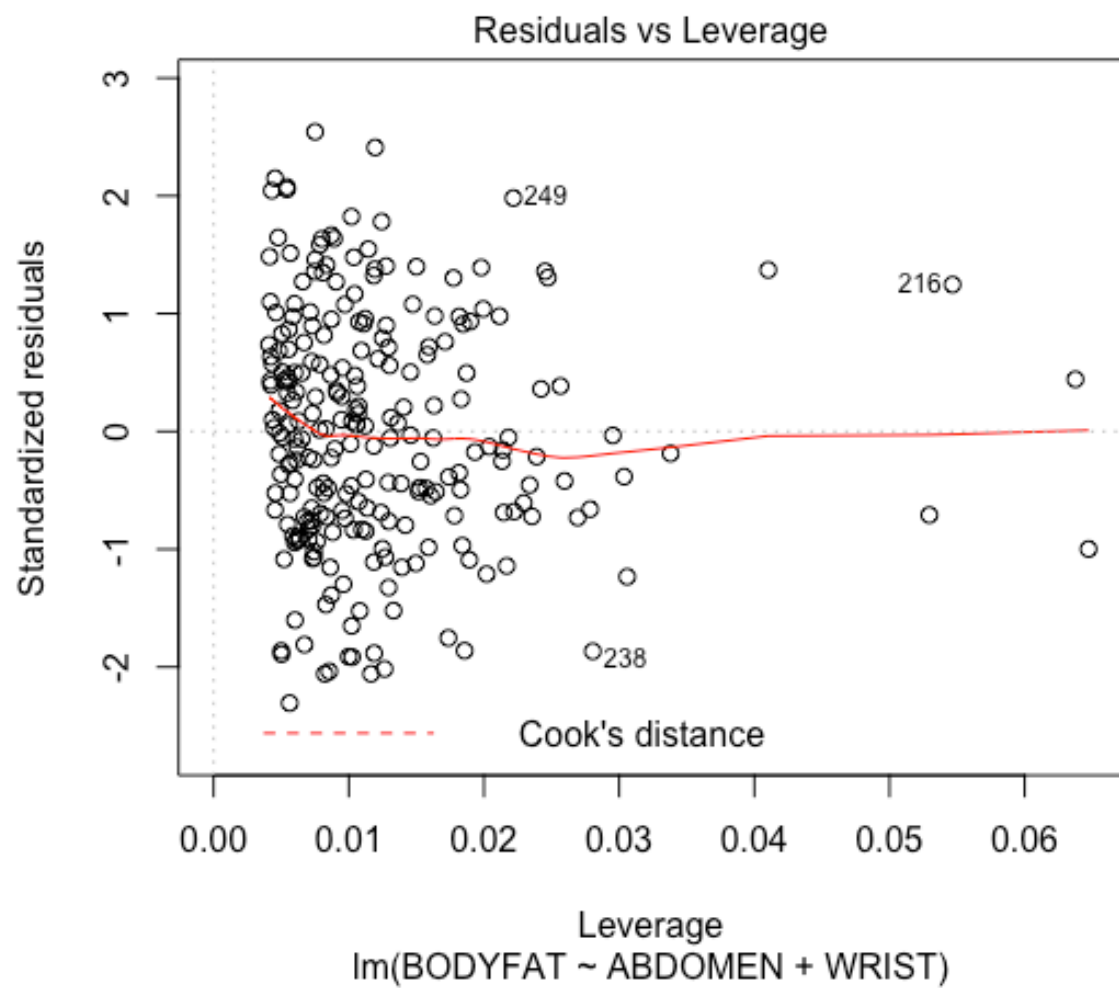
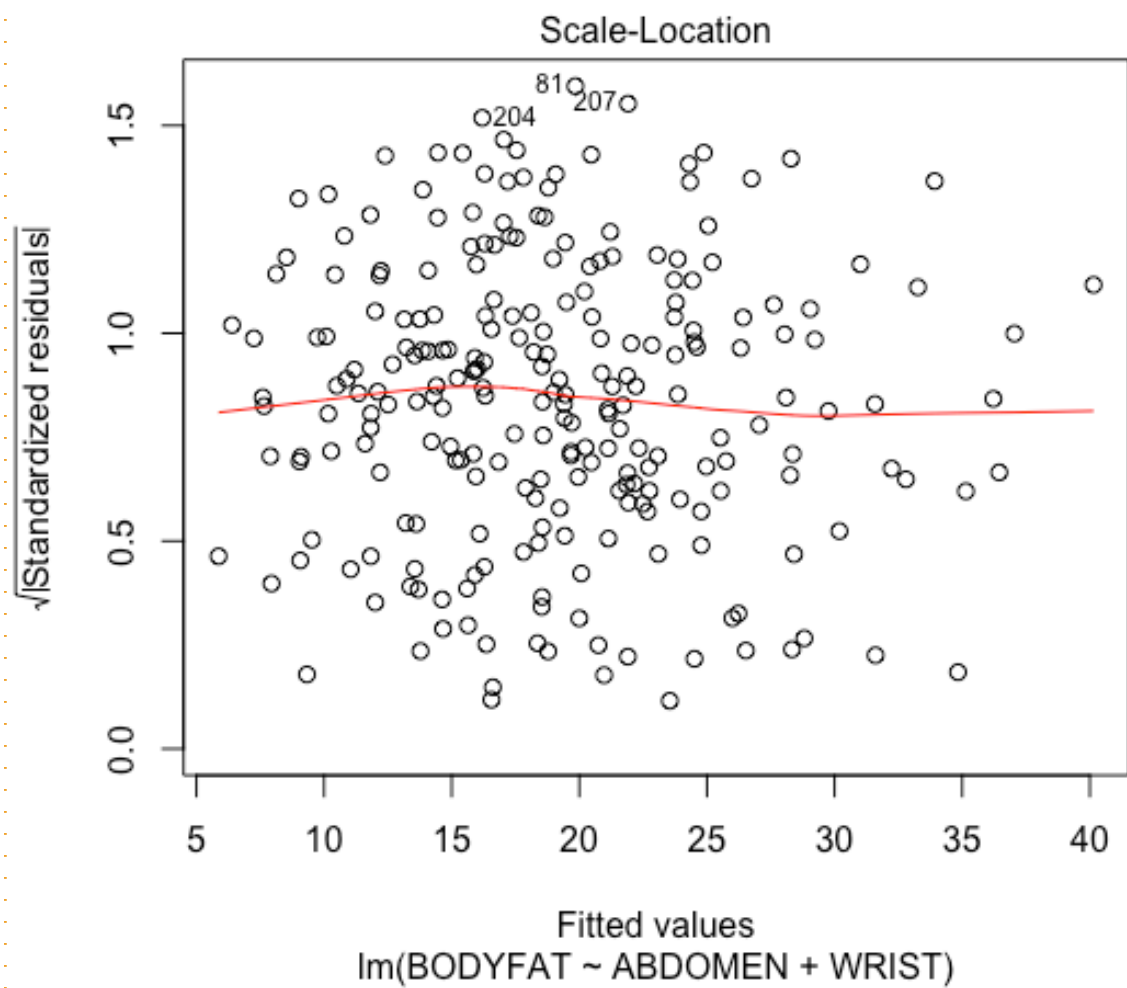
RULE OF THUMB

$$\text{BODYFAT} = 0.73 * \text{ABDOMEN} - 2.03 * \text{WRIST} - 11.2$$

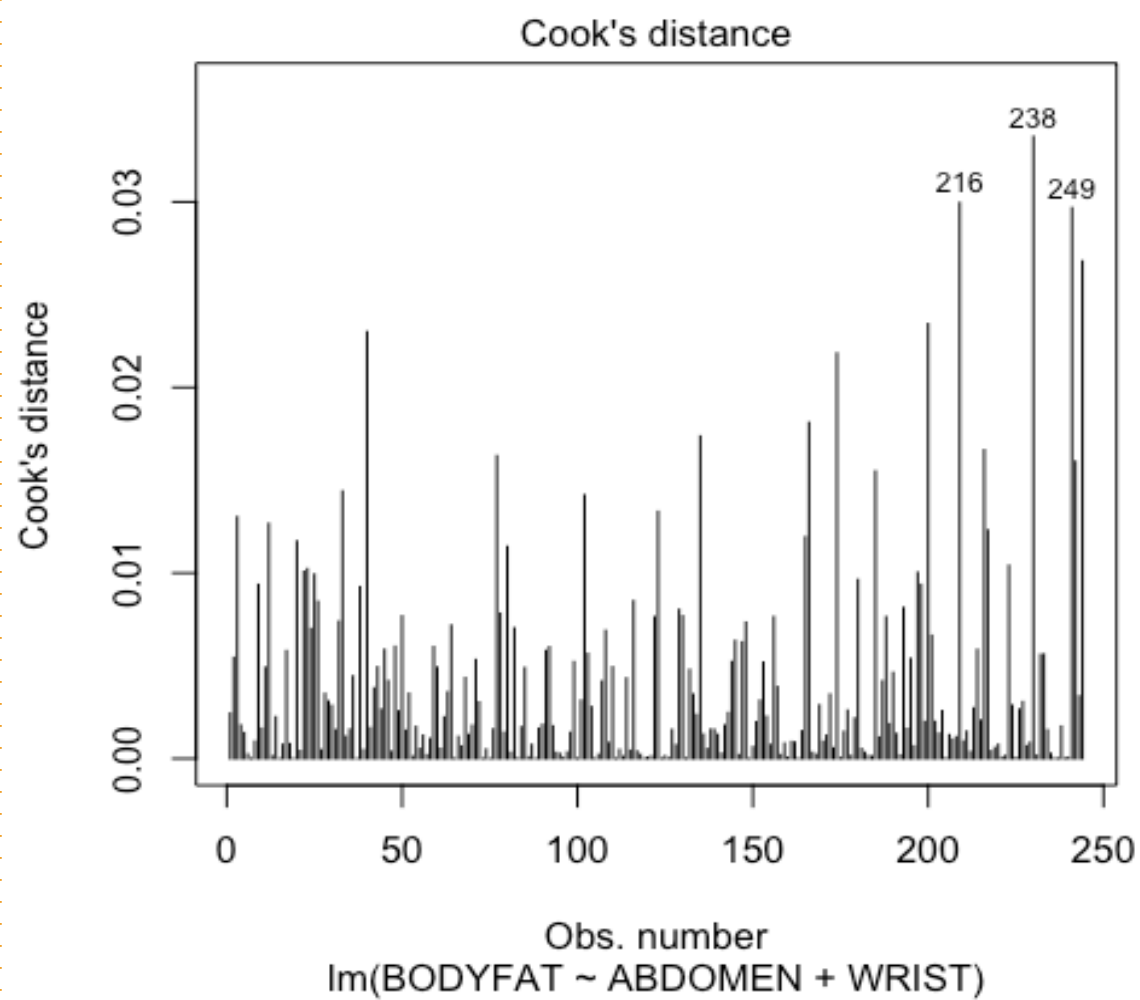
Diagnostics



Diagnostics



Diagnostics



PROs and Cons

- Our final model is simple and straightforward. The variables included for prediction is easy to get.
- The model is robust.
- The model is accurate as the adjusted R-squared is greater than 0.6.
- We only consider two variables and ignored the interaction between these variables.
- Our data sample only contains 252 observations, if the weight is too high or too low, the prediction may not be so accurate.

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. In the center, there is a large orange speech bubble with a white border. The text "THANK YOU FOR WATCHING" is written in white, bold, uppercase letters inside the speech bubble.

THANK YOU FOR WATCHING