Bodyfat Prediction

Yilun Chen
Jiyun Chen
Huiyu Jiang

University of Wisconsin-Madison

February 7, 2018
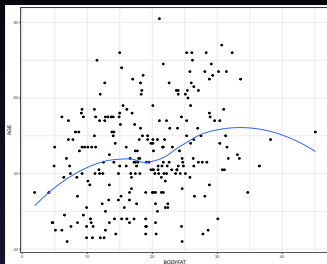
Outline

- Background and Data Description
- Model Selection and Data Processing
- Variable Selection
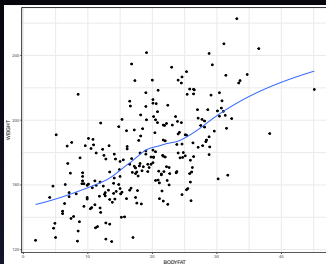- Model Diagnostic and Prediction

Background

- Motivation:
  Use data to train models to make predictions.
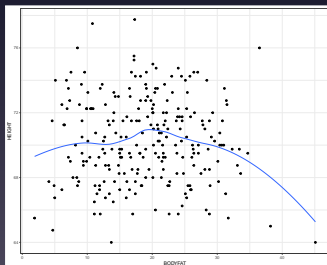- Target:
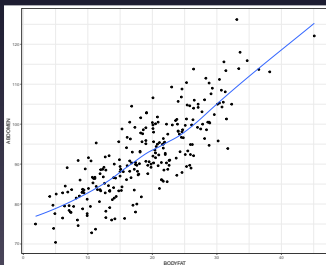 A clean and robust model

Model Selection

(a) age

(b) weight

(c) height

(d) abdomen

Data Processing

- false record:

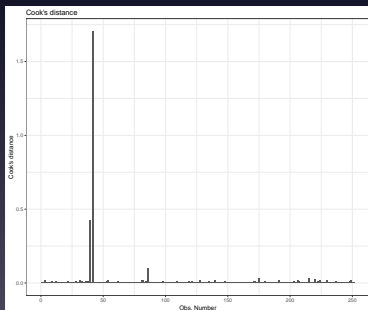|	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
182	0	40	118.5	68	18.1

- check outliers:

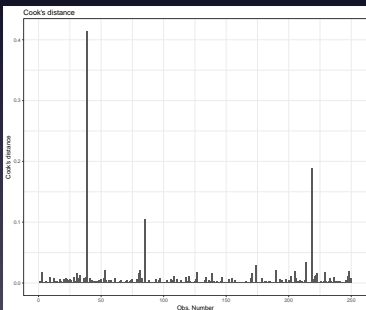|	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
42	31.7	44	205	29.5	29.9

(e) 42

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
39	33.8	46	363.15	72.25	48.9

(f) 39

(g)

(h)

Variable Selection

- AIC Stepwise:

```
Call:
lm(formula = BODYFAT ~ ABDOMEN + WEIGHT + WRIST + BICEPS + AGE,
    data = data_body)

Residuals:
    Min      1Q  Median      3Q     Max
-9.4993 -2.6962 -0.2706  3.1972  8.7266

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.71669    7.06737  -3.356 0.000918 ***
ABDOMEN       0.83796    0.06314  13.272  < 2e-16 ***
WEIGHT       -0.08539    0.03055  -2.795 0.005607 **
WRIST        -1.65891    0.45052  -3.682 0.000285 ***
BICEPS        0.27709    0.13981   1.982 0.048619 *
AGE           0.03881    0.02615   1.484 0.139047
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.95 on 243 degrees of freedom
Multiple R-squared:  0.7356,    Adjusted R-squared:  0.7301
F-statistic: 135.2 on 5 and 243 DF,  p-value: < 2.2e-16
```
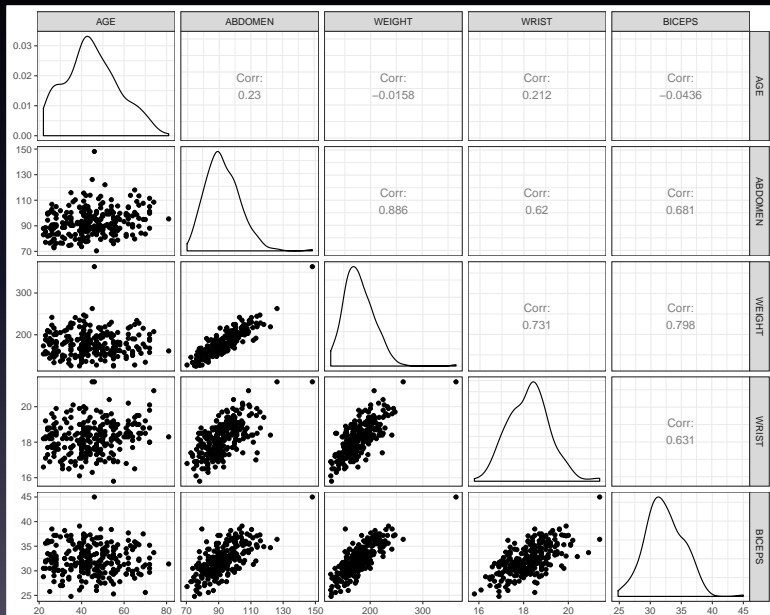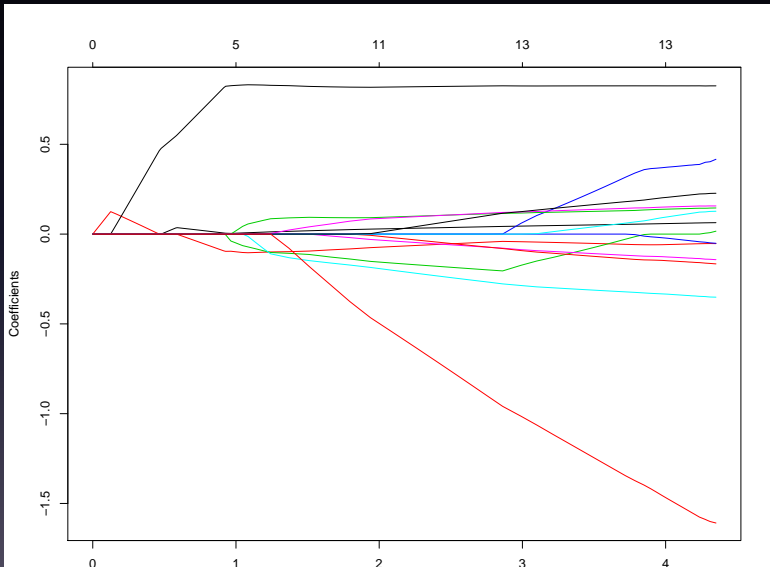
Variable Selection

Variable Selection

- elastic net:

Variable Selection

WEIGHT	4.1483
ABDOMEN	4.1483

```
Call:
lm(formula = BODYFAT ~ WEIGHT + ABDOMEN, data = df)

Residuals:
     Min      1Q  Median      3Q     Max
-10.0573 -2.9614 -0.1035  2.9452  9.9097

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.28955    2.45945  -17.20  < 2e-16 ***
WEIGHT       -0.12314    0.01955   -6.30 1.36e-09 ***
ABDOMEN       0.90011    0.05179   17.38  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.054 on 246 degrees of freedom
Multiple R-squared: 0.718,    Adjusted R-squared: 0.7157
F-statistic: 313.2 on 2 and 246 DF,  p-value: < 2.2e-16
```
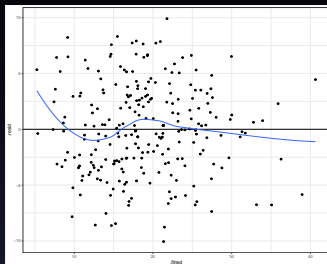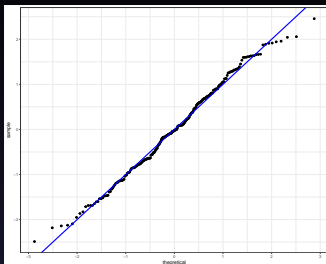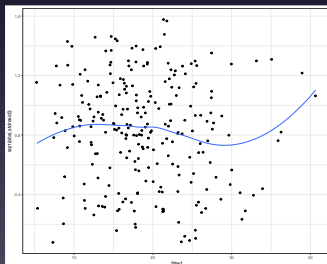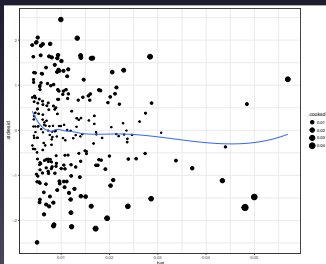
Diagnostic

(i) residule

(j) QQ-plot

(k) std-res

(l) cook distance

Prediction

Confidence intervals for parameters