Bodyfat Analysis

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

- Thesis Statement:

  An analysis of the male bodyfat database:

  Using a linear model to infer and predict the male bodyfat based on three factors.

- Target:

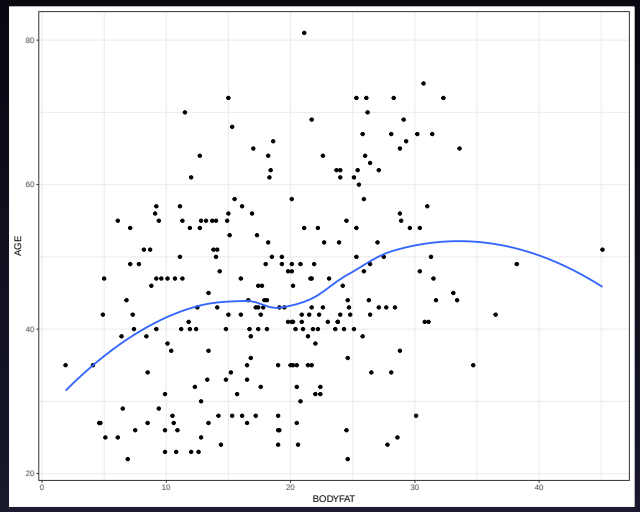
  A simple, convenient, and robust model.

- Data Description:

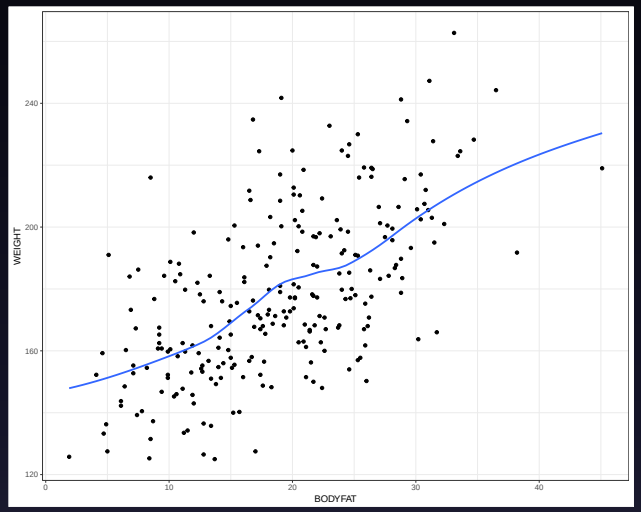
IDNO	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN
1	12.6	1.0708	23	154.25	67.75	23.7	36.2	93.1	85.2
2	6.9	1.0853	22	173.25	72.25	23.4	38.5	93.6	83.0
3	24.6	1.0414	22	154.00	66.25	24.7	34.0	95.8	87.9
4	10.9	1.0751	26	184.75	72.25	24.9	37.4	101.8	86.4
5	27.8	1.0340	24	184.25	71.25	25.6	34.4	97.3	100.0
6	20.6	1.0502	24	210.25	74.75	26.5	39.0	104.5	94.4

Model Selection

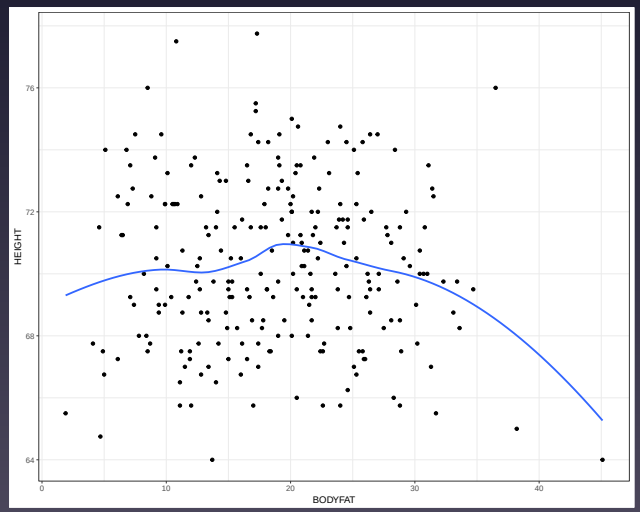
Is the linear model good enough to catch the trend on bodyfat?



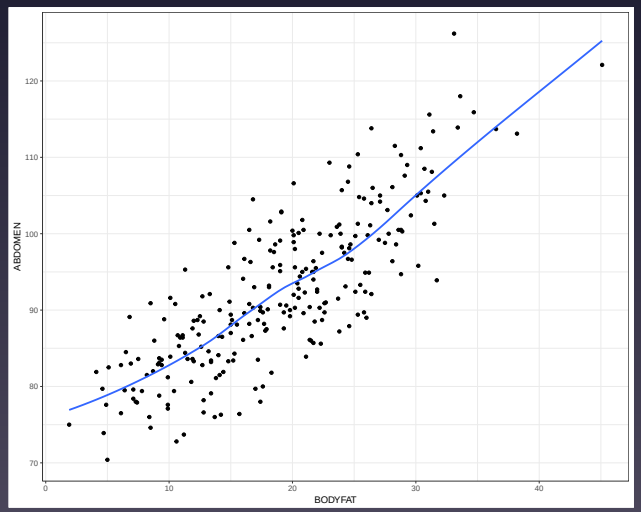
(a) age



(b) weight



(c) height



(d) abdomen

Data Processing

- false record:

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
182	0	40	118.5	68	18.1

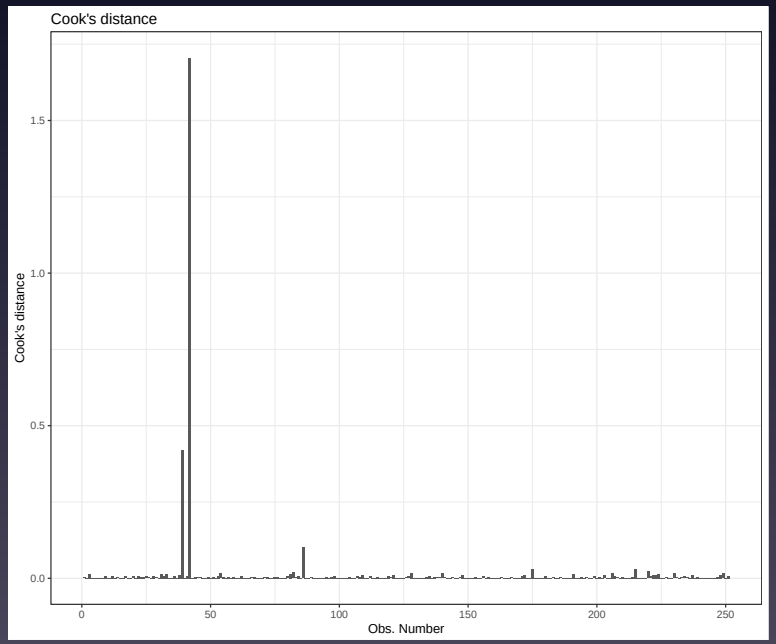
- check outliers:

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
42	31.7	44	205	29.5	29.9

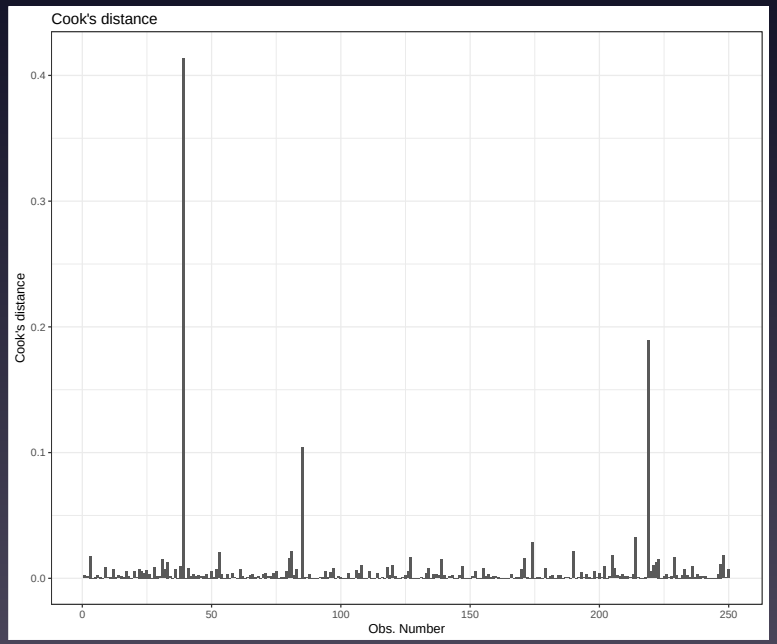
(e) 42

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
39	33.8	46	363.15	72.25	48.9

(f) 39



(g)



(h)

Variable Selection

Select the variables based on stepwise method and elastic net.

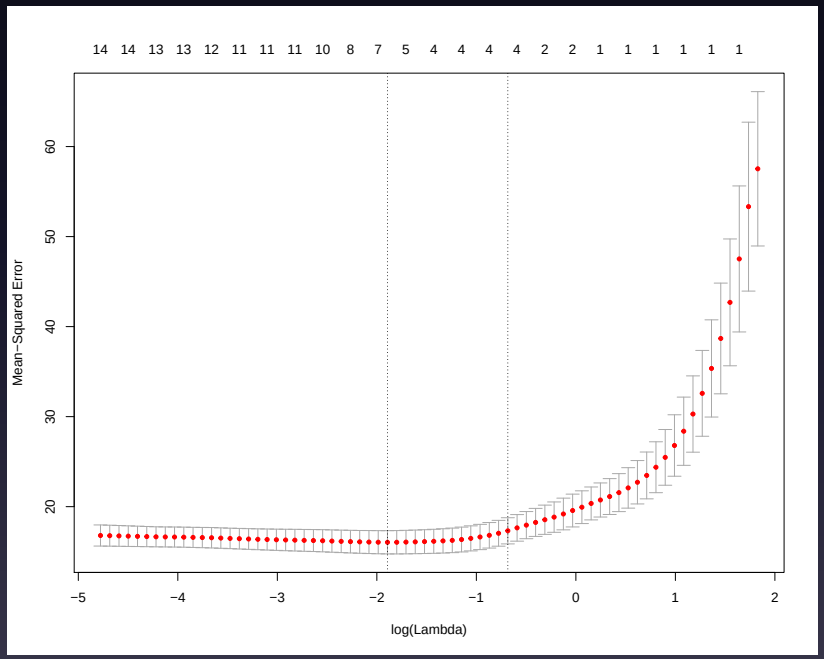
Stepwise:

$$lnL(\mu, \sigma^2) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

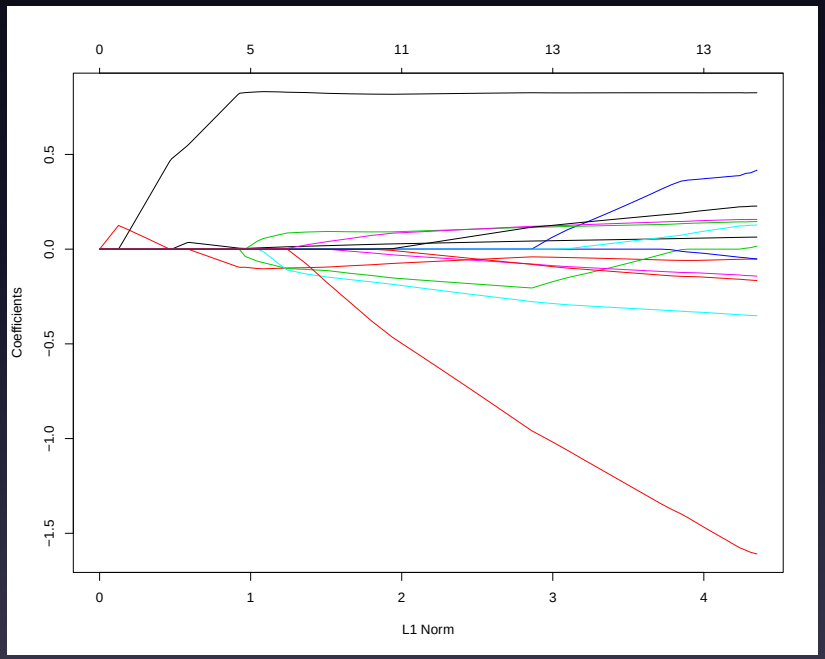
- AIC: $2k - 2ln(\hat{L})$
  ABDOMEN WEIGHT WRIST BICEPS AGE
- BIC: $ln(n)k - 2ln(\hat{L})$
  ABDOMEN WEIGHT WRIST

Variable Selection

- elastic net: $\hat{\beta} = argmin_{\beta}(||y - X\beta||^2 + \lambda_2||\beta||^2 + \lambda_1||\beta||_1)$



(i) cv



(j) regulization path

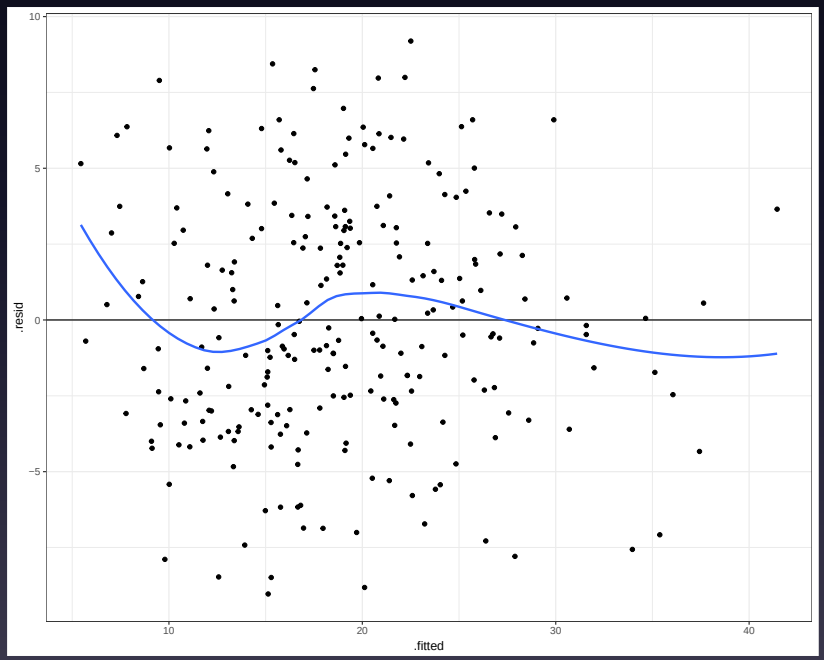
Variable Selection

WEIGHT	5.6178
ABDOMEN	4.1857
WRIST	2.0988

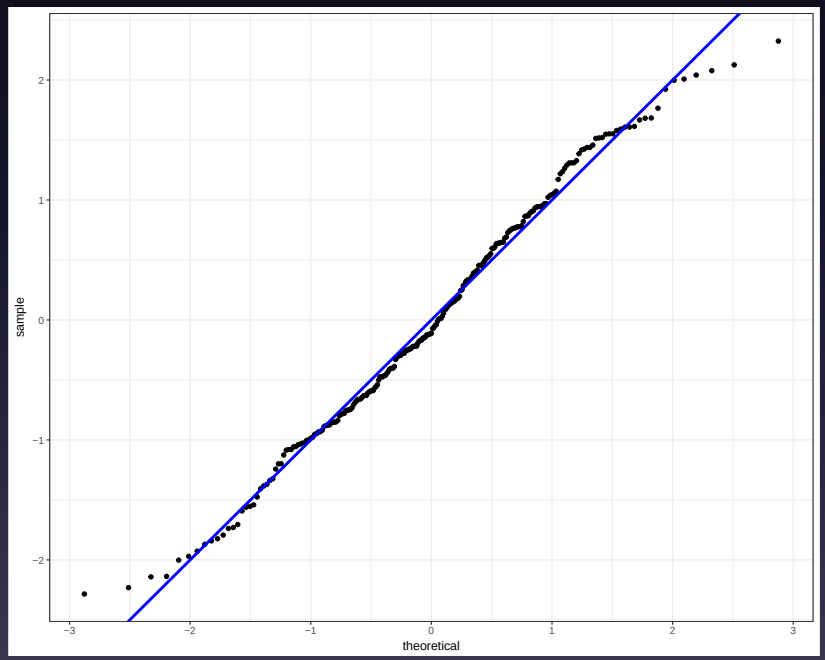
```
Call:
lm(formula = BODYFAT ~ WEIGHT + ABDOMEN + WRIST, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-9.0322 -2.9552 -0.4383  3.0110  9.1926

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.99356    6.23603  -3.848 0.000152 ***
WEIGHT       -0.08678    0.02234  -3.885 0.000132 ***
ABDOMEN       0.88476    0.05108  17.320  < 2e-16 ***
WRIST        -1.28161    0.40274  -3.182 0.001650 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.981 on 245 degrees of freedom
Multiple R-squared:  0.7292,    Adjusted R-squared:  0.7259
F-statistic: 219.9 on 3 and 245 DF,  p-value: < 2.2e-16
```

Diagnostic

Rule of Thumb:
Bodyfat = -24 - 0.1*Weight(lb) + 0.9*Abdomen(cm)
- 1.3*Wrist(cm)

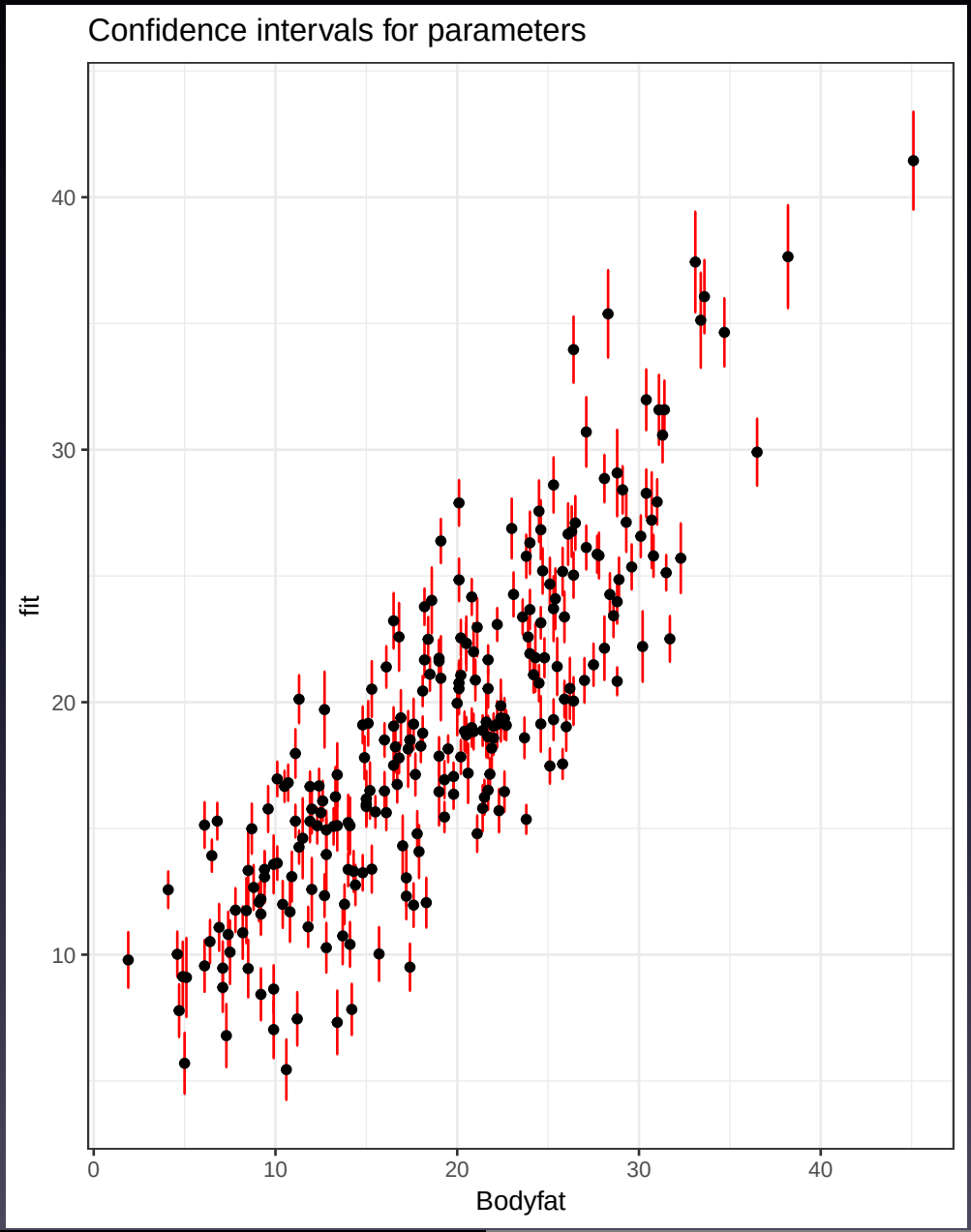


(k) residual



(l) QQ-plot

Prediction


Confidence intervals for parameters
fit
Bodyfat
0
10
20
30
40
10
20
30
40

Summary

- Strength:
  1.easy to implement
  2.not need a lot of information
- Weakness:
  rough estimation (relatively large residules)
- Illustrative example:
  Weight: 136.75lb, Abdomen: 77.0cm, Wrist: 16.5cm
  Estimated Bodyfat: 10.15
  Confidence Interval: (9.05,11.25)