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
- Summary and Reference

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.

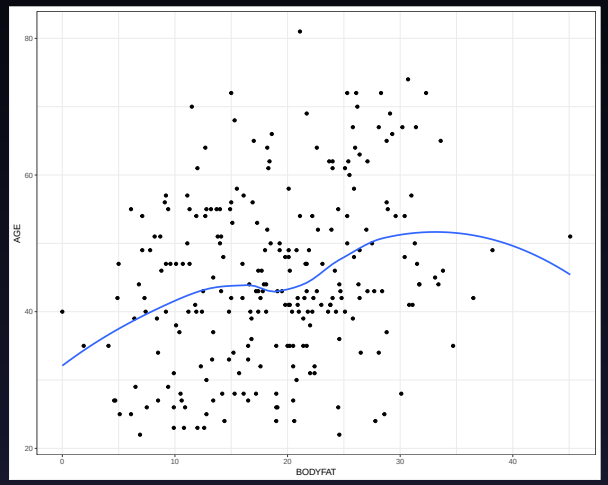
- Data Description:

BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST
12.6	1.0708	23	154.25	67.75	23.7	36.2	93.1
6.9	1.0853	22	173.25	72.25	23.4	38.5	93.6
24.6	1.0414	22	154.00	66.25	24.7	34.0	95.8

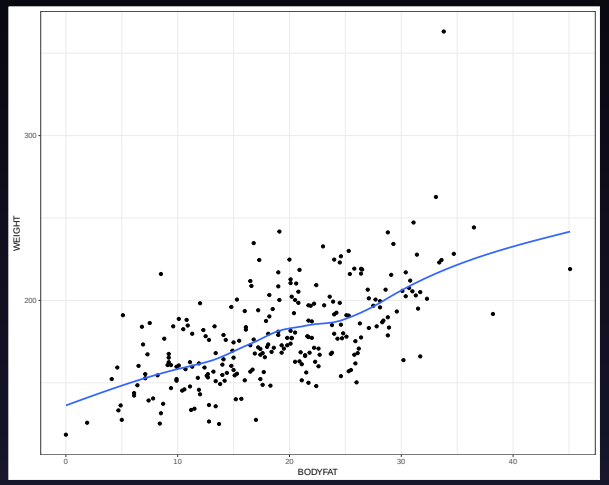
ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6

Model Selection

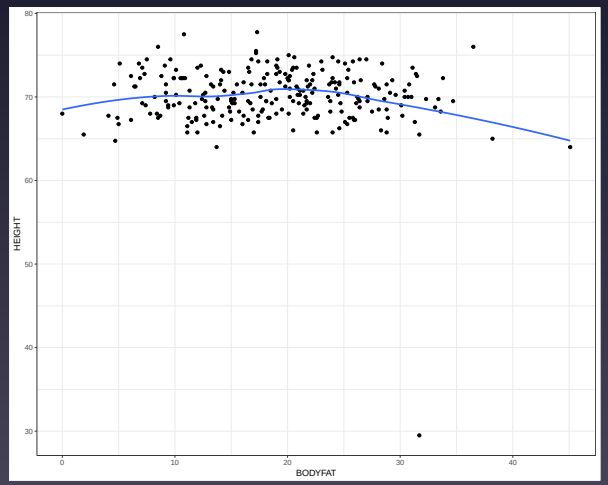
Is the linear model good enough to catch the trend on bodyfat?



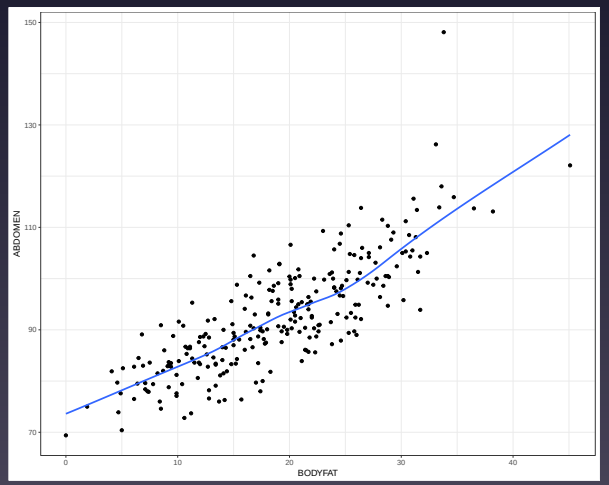
(a) age



(b) weight



(c) height



(d) abdomen

Data Processing

- false record:

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
182	0	40	118.5	68	18.1

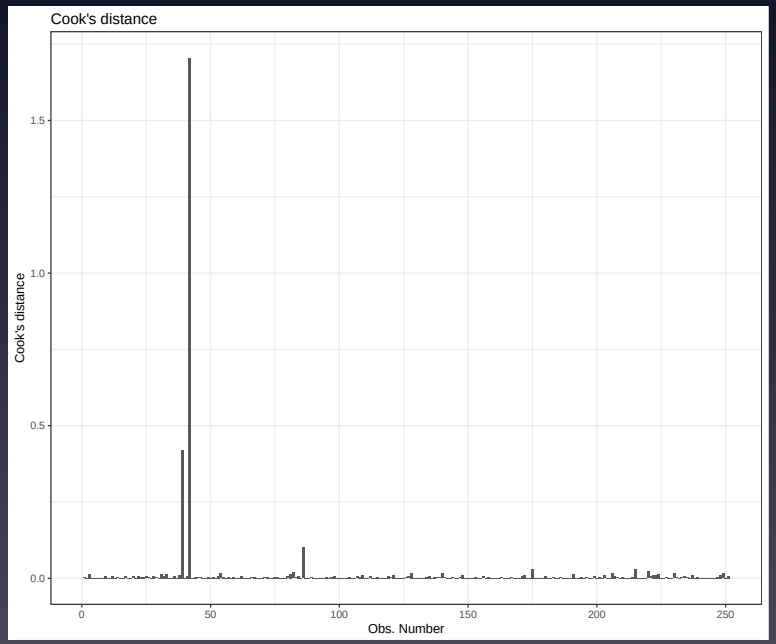
- check outliers:

	BODYFAT	AGE	WEIGHT	HEIGHT	ADIPOSITY
42	31.7	44	205	29.5	29.9

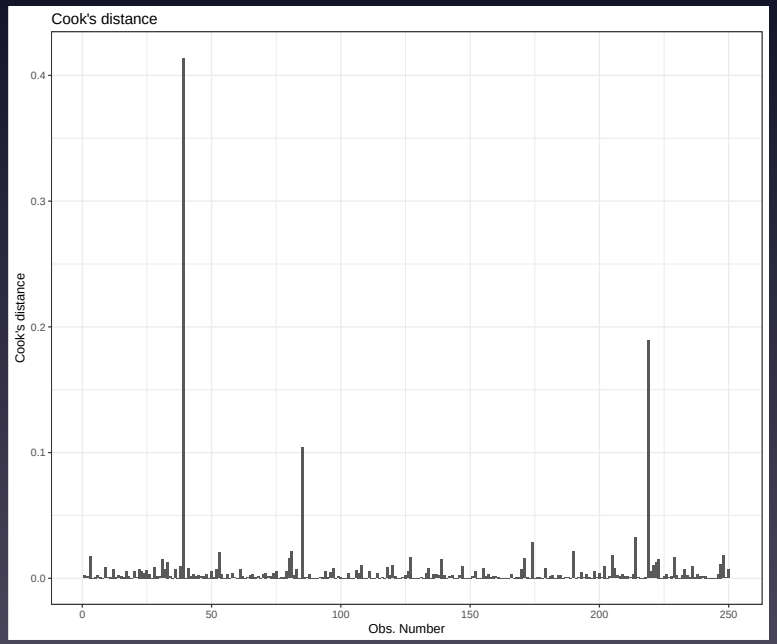
(e) 42

	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.

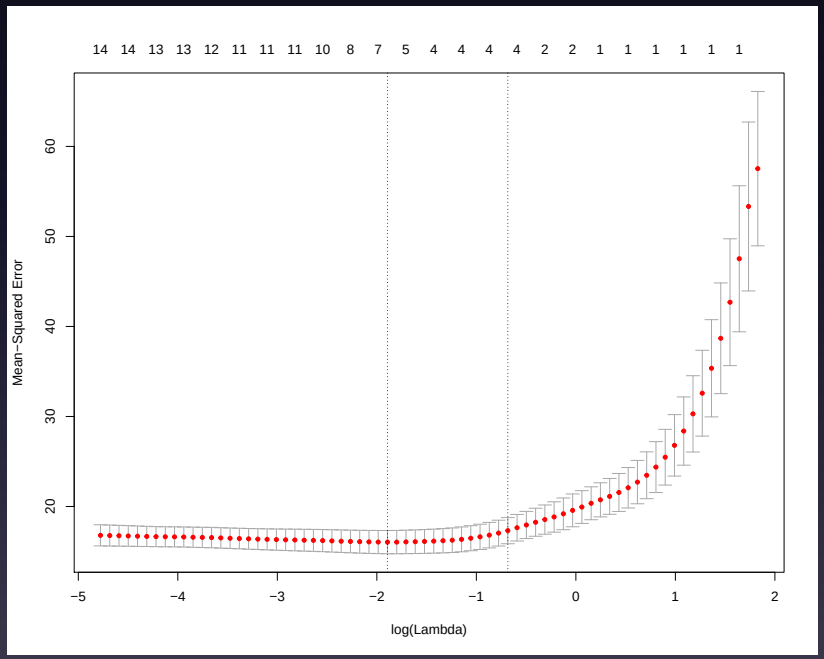
Stepwise:

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

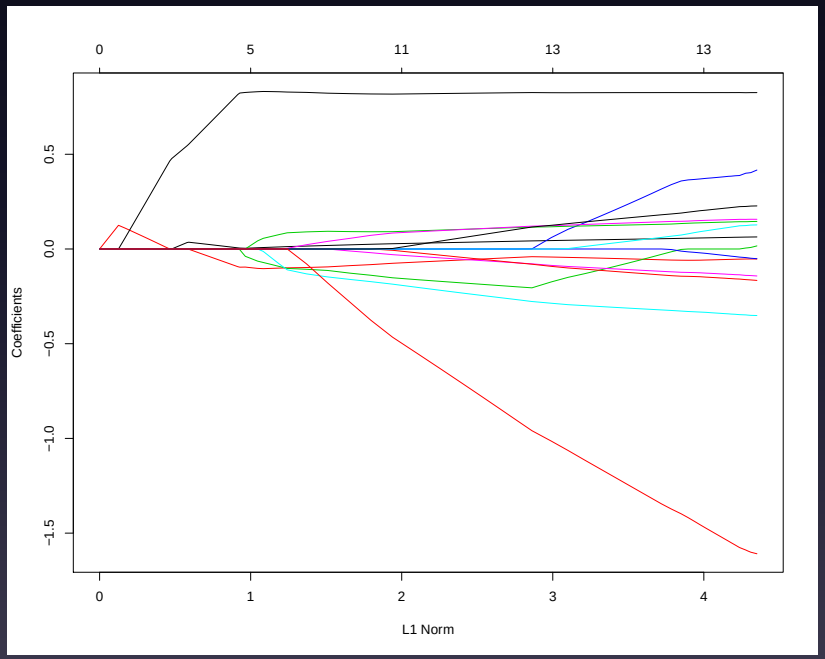
- AIC: $2k - 2ln(\hat{L})$
  ABDOMEN WEIGHT WRIST BICEPS AGE
- BIC: $ln(n)k - 2ln(\hat{L})$
  ABDOMEN WEIGHT WRIST

Variable Selection

- elastic net:
  $\hat{\beta} = argmin_{\beta}(||y - X\beta||^2 + \lambda(1-\alpha)||\beta||^2 + \lambda\alpha||\beta||_1)$



(i) cv



(j) regulization path

Final Model

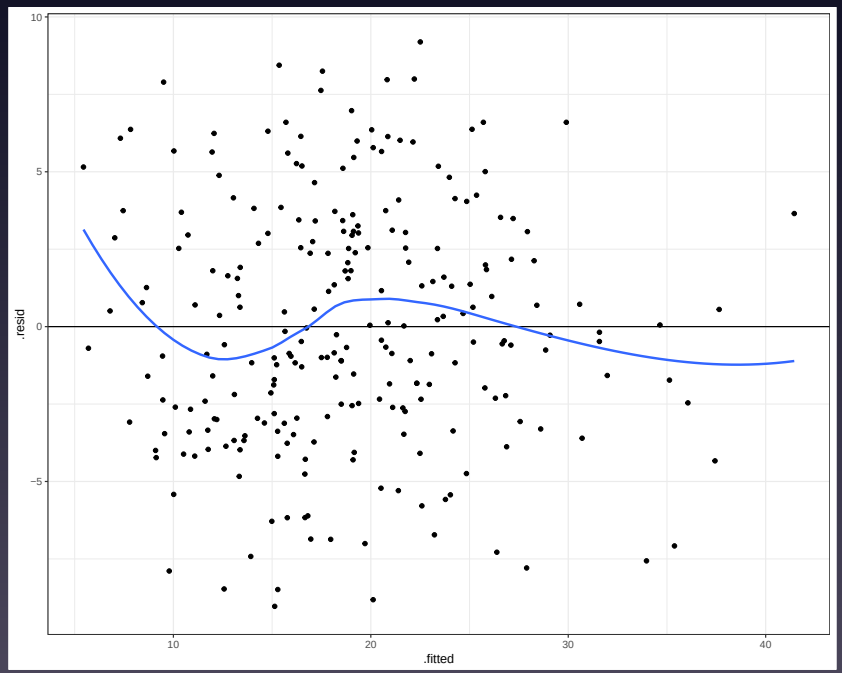
- Variables:
  WEIGHT ABDOMEN WRIST
- trained model:
  Bodyfat = -23.994 - 0.087*Weight(lb) +
  0.885*Abdomen(cm) - 1.282*Wrist(cm)
- Residual standard error: 3.981
- R^2: 0.7292
- p-value: < 2.2e-16
- Rule of Thumb:
  Bodyfat = -24 - 0.1*Weight(lb) + 0.9*Abdomen(cm)
  - 1.3*Wrist(cm)

Diagnostic

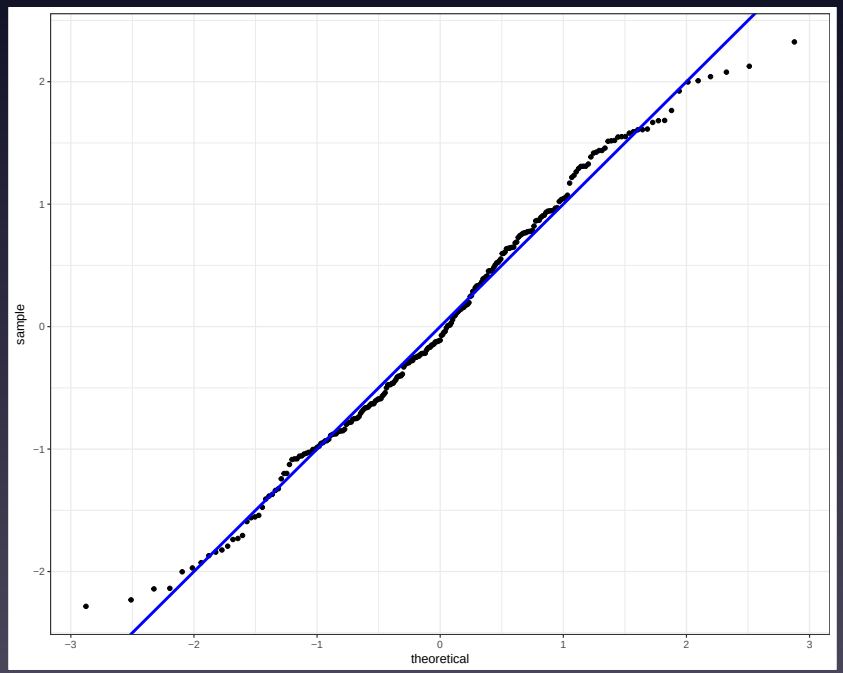
- VIF:

Variables	VIF
Weight	5.6178
Abdomen	4.1857
Wrist	2.0988

- Residuals:

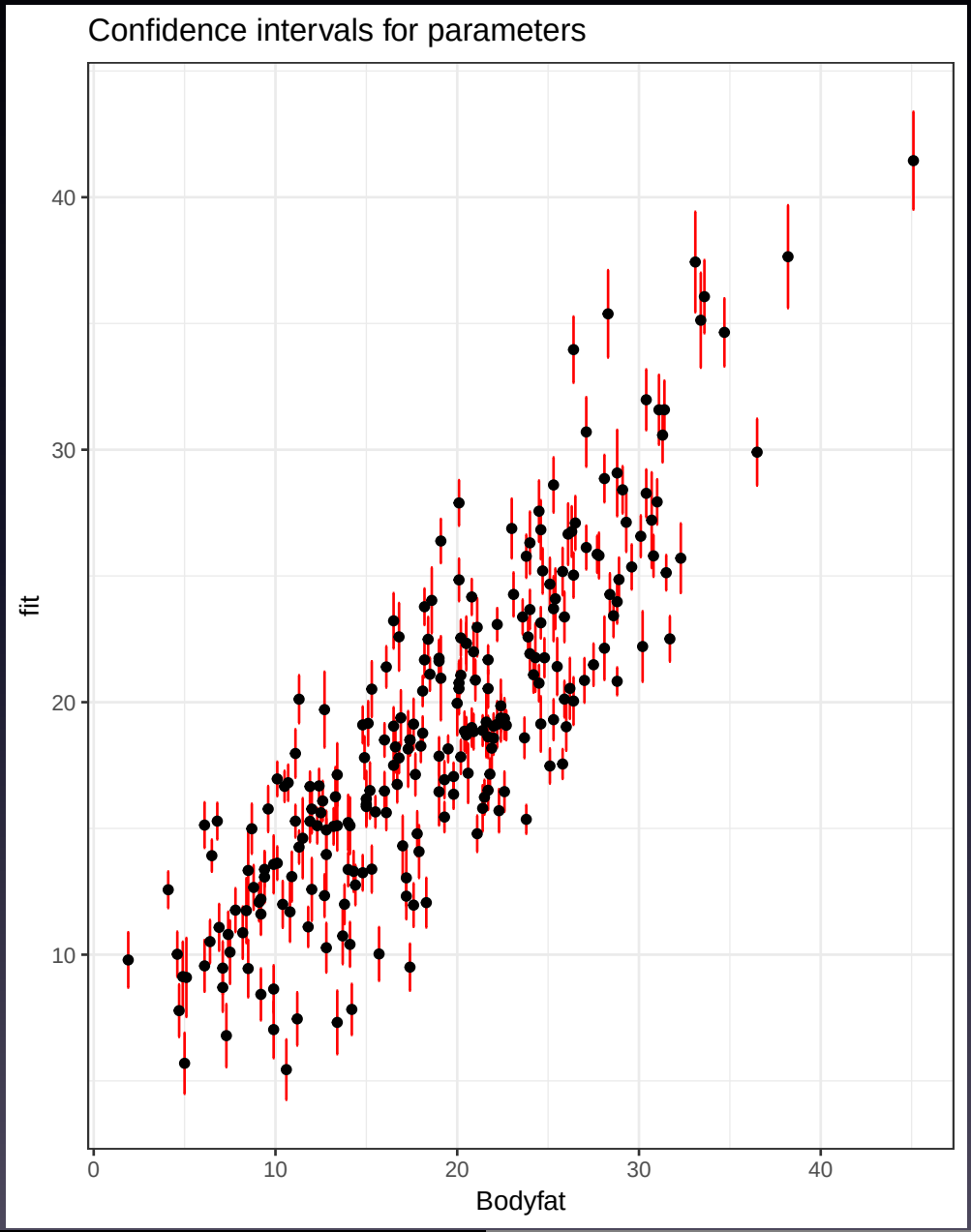


(k) residual



(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:
  lack of accuracy (relatively large residuals)
- Illustrative example:
  Weight: 136.75lb, Abdomen: 77.0cm, Wrist: 16.5cm
  Estimated Bodyfat: 10.15
  Confidence Interval: (9.05,11.25)

Appendix

References:

- Burnham, K. P., Anderson, D.R. (2004),*Multimodel inference: understanding AIC and BIC in Model Selection*, Sociological Methods & Research.
- Cook, R. Dennis (1977), *Detection of influential Observations in Linear Regression*, American Statistical Association.
- Friedman J, Hastie T, Tibshirani R., (2009), *Regularization Paths for Generalized Linear Models via Coordinate Descent*, Journal of Statistical Software.
- Roger W. Johnson (1996), *Fitting Percentage of Body Fat to Simple Body Measurements*, Journal of Statistics Education.