Method of body fat estimation

Chenlai Shi, Guanxu Su, Jing Tao

UW-Madison

February 3, 2018

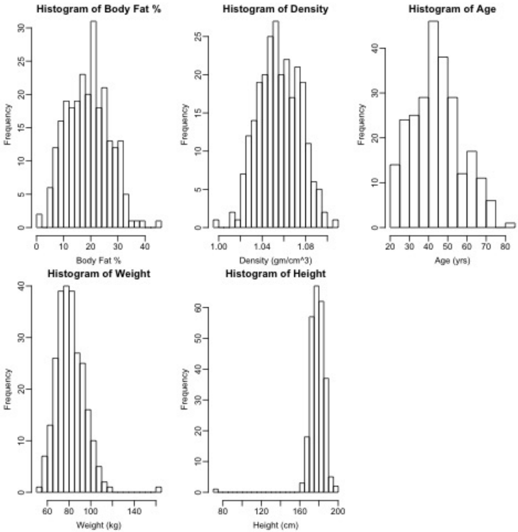
Motivation and introduction

- Wikipedia's definition of body fat percentage is the total mass of fat divided by total body mass, times 100.
- Measure of fitness level based on various theoretical approaches about the relationships between body fat percentage, health, athletic capacity, etc.
- Different measurement techniques, such as underwater weighing, whole-body air displacement plethysmography, body average density measurement
- New method,is easy-to-use, robust, scalable and accurate.

rule-of-thumb method

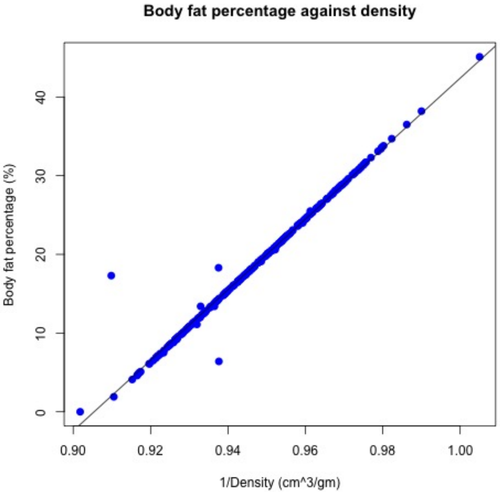
- Bodyfat $= (-5500 + 6\text{Weight} + 1000\text{Hip} - 1000\text{Wrist} + 3\text{Abdomen}^3 - 3.5\text{Hip}^3)/\text{Weight}$
- 6 times weight(kg) plus 1000 times hip circumference(dm) minus 1000 times wrist(dm) plus 3 times abdomen 2 circumference(dm) cube minus 3.5 times hip circumference(dm) cube minus 5500 over weight(kg).
- For example, a man of weight 70 kg, hip circumference 10 dm, abdomen 9 dm, wrist 1.7 dm, estimated body fat is about 27%

Data clean



Histogram of Body Fat %
Frequency
Body Fat %
Histogram of Density
Frequency
Density (gm/cm^3)
Histogram of Age
Frequency
Age (yrs)
Histogram of Weight
Frequency
Weight (kg)
Histogram of Height
Frequency
Height (cm)

Data clean



Body fat percentage against density
Body fat percentage (%)
0
10
20
30
40
0.90
0.92
0.94
0.96
0.98
1.00
1/Density (cm^3/gm)

Model

What we know

- Relationship between *Bodyfat* and $\frac{1}{Density}$ is nearly perfect
- $\frac{1}{Density} = \frac{Volume}{Weight}$
- *Volume* has closer relationship with our measurable quantitives such as *Weight*, *Hip*, *Wrist*, etc.
- Measurable variables $\Rightarrow$ Volume $\Rightarrow \frac{1}{\text{Density}} \Rightarrow$ Bodyfat

Model

- Model 1

$$\begin{aligned} \text{Volume}_i =& \beta_{10} + \beta_{11}\text{Weight}_i + \beta_{12}\text{Hip}_i + \beta_{13}\text{Wrist}_i \\ &+ \beta_{14}\text{Abdomen}_i^3 + \beta_{15}\text{Hip}_i^3 + \varepsilon_{1i}, \quad \varepsilon_i \sim N(0, \sigma_2^2) \end{aligned}$$

- Model 2

$$\text{Bodyfat}_i = \beta_{20} + \beta_{21}\frac{1}{\text{Density}_i} + \epsilon_2 i, \quad \epsilon_i \sim N(0, \sigma_2^2)$$

- Final Model

$$\hat{\text{Bodyfat}} = \beta_{20} + \beta_{21}\frac{Volume}{\text{Weight}}$$

Fitted model

- $\hat{\text{Volume}} = -1.206 + 0.920\text{Weight} + 0.217\text{Hip} - 0.250\text{Wrist} + 6.105(\text{Abdomen}^3/10^6) - 7.604(\text{Hip}^3/10^6)$
- Weight, Height, Hip circumference, Wrist circumference, abdomen 2 circumference explains about 99.7% of the variation in Volume.
- Men gain $9.20dm^3$ on average for every extra kilogram weight with other variables being the same.
- Men gain $0.217dm^3$ on average for every extra centimetre hip circumference with other variables being the same.

Fitted model 1

- Men loss $0.250 dm^3$ on average for every extra centimeter wrist circumference with other variables being the same.
- Men gain $6.105 \times 10^{-6} dm^3$ on average for every extra cm^3 abdomen 2 circumference^3 with other variables being the same.
- Men loss $7.604 \times 10^{-6} dm^3$ on average for every extra cm^3 hip^3 with other variables being the same.

Fitted model 2

- Bodyfat $= -414 + \frac{457}{\text{Density}}$
- $\frac{1}{Density}$ explains about 99.98% of the variation in Volume (based on R^2).
- Men gain 457% of bodyfat% on average for every extra unit $\frac{1}{Density}$.
- We can almost surely declare that there is linear relationship between $\frac{1}{Density}$ and *Bodyfat*.

Final Model

- $$\begin{aligned}\text{Bodyfat} =& -414 + (-5511 + 420\text{Weight} + 99\text{Hip} - 114\text{Wrist} \\ &+ 2790(\text{Abdomen}^3/10^6) - 3475(\text{Hip}^3/10^6))/\text{Weight}\end{aligned}$$

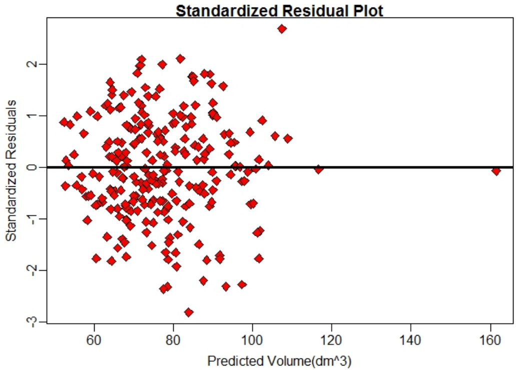
- The equaltion can be approximately expressed by:

$$\begin{aligned}\text{Bodyfat} =&(-5500 + 6\text{Weight} + 100\text{Hip} - 100\text{Wrist} \\ &+ 2800(\text{Abdomen}^3/10^6) - 3500(\text{Hip}^3/10^6))/\text{Weight}\end{aligned}$$

K-fold cross validation

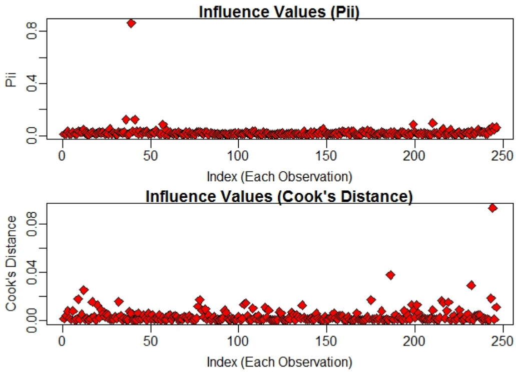
- Dataset is Divided into 10 groups randomly
- Choose 1 group as testing dataset and one by one untill all the group is chosen
- Get ten MSE in all and the average of them is about 14

Diagnostics



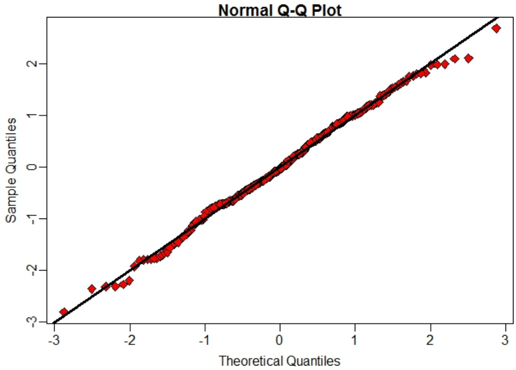
The standard errors are symmetrically distributed around 0, homoskedasticity assumption holds.

Diagnostics



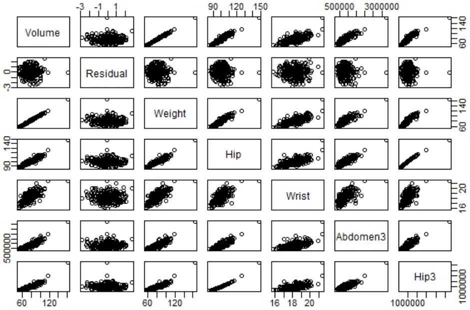
Observation No.39 has extremely high Pii but does not has a high cook's distance. Hence, the observation will not be excluded

Diagnostics



Normality assumption holds

Diagnostics



All the variables have linear relationship with volume and are independent from residual

Strength and weakness

- Linearity: All the explanatory variables have strong linear relationship with the response variables in both two models.
- Additivity: In the volume model, most variables can be added to each other.
- Constant effects: since we don't split the whole dataset into different groups, our model may not be very precise for specific groups.
- Fixed X: This seems reasonable for this data.
- Normally distributed errors: after our data cleaning procedure, the QQ plot of the model looks perfect.
- Constant variance: satisfied by looking at the residual plot.
- The prediction of each model of the two steps are both very precise.
- The final model is complicated in calculation.

Thank You!