BodyFat Analysis

GROUP 16 Jinchen Gong, Lanxi Zhang, Yiyuan Li

Introduction and Motivation

- Body fat percentage is the percentage of total body weight that is fat multiplied by 100. it includes both essential and stored fat, which is vital for life and reproduction, and is an indicator of the fitness of an individual's body composition.
- Measuring body fat accurately can be expensive, so people want to find simpler ways to estimate body fat. Our team's goal was to develop an inexpensive, efficient, reliable and accurate method using readily available clinical data.
- We find that a multiple linear model can predicts BODYFAT very accurately and remains simple and robust.

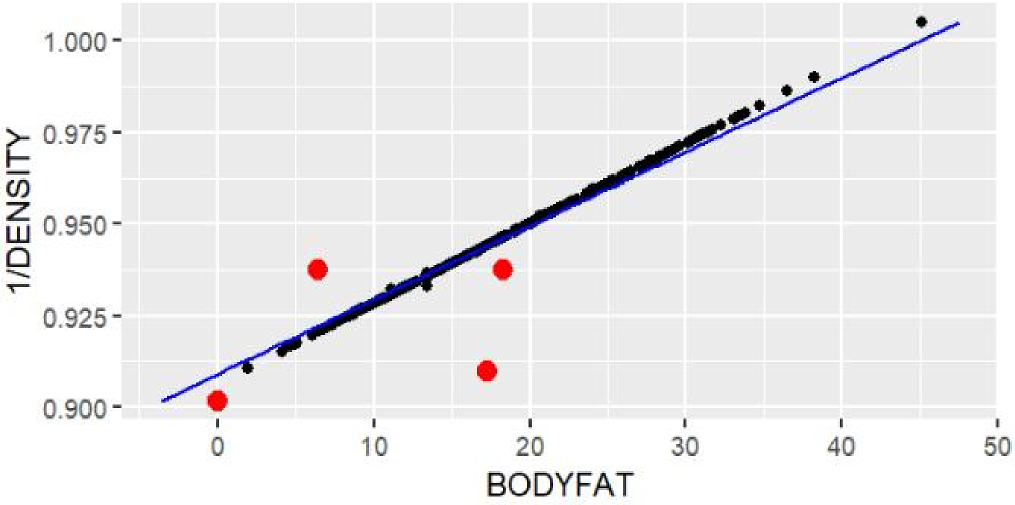
Background Information

- "Siri's equation":

$$100 \times \mathrm{BODYFAT} = \frac{495}{\mathrm{DENSITY}} - 450$$

Data Cleaning

1. Using a linear model fit to find data points deviating significantly: Outliers(marked in red)



2. 42nd data point's Height 29.5 inches (75cm) ——unusually short: might be a typo

So we remove these points.

Modeling

- Goal: achieve a simple, robust, and accurate model
- Use stepwise model selection: **AIC** and **BIC** standards
- Prevent overfitting: **Lasso regression**
- Fit multiple linear models based on the selected variables

Model Comparing

METHOD	R.Squared	MSE.test	VIF	Sim
AIC	0.7441	17.3468	>10	9
BIC	0.7320	15.5411	<10	4
LASSO	0.7416	17.1149		4

Model Comparing

- 1. **Accuracy**: R ^ 2 of the three models are very close, but the MSE of the BIC selection model is significantly smaller than the other two models, indicating that the BIC model has smaller testing errors and better predictive performance.
- 2. **Simplicity**: only four variables in the BIC and Lasso models are smaller than those in the AIC model, indicating that the BIC and Lasso models have lower complexity and are simpler and more convenient.
- 3. **Multicollinearity**: the VIF of AIC is greater than 10, indicating that the correlation between model independent variables may be higher. Therefore, we have decided to choose the BIC model.
- Therefore, we have decided to choose the BIC model.

Final Model

$$BodyFat\% = -34.91-0.35\times WEIGHT+1.00\times ABDOMEN+0.50\times FOREARM-1.37\times WRIST$$

Analysis:

1. A man weighing approximately 85 kg, with a waist circumference of 92 cm, a forearm circumference of 31 cm, and a wrist circumference of 20 cm would have a predicted body fat percentage of 16.4%.

2. The 95% prediction interval for his body fat percentage lies between 14.9% and 17.9%.

3. Body fat increases by nearly 10% for every 10 cm increase in other measurements constant.

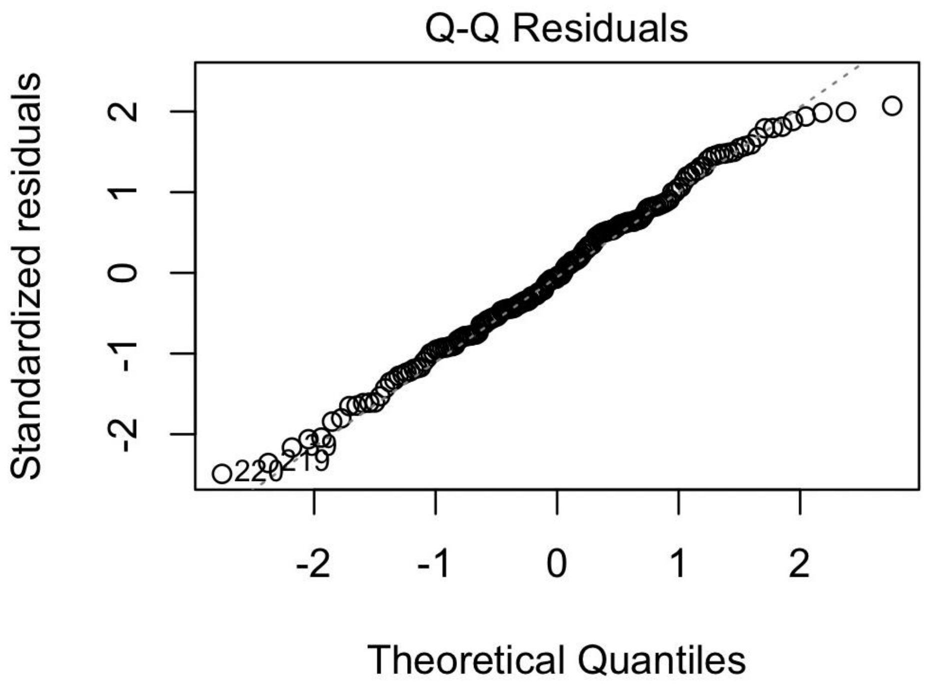
Final Model

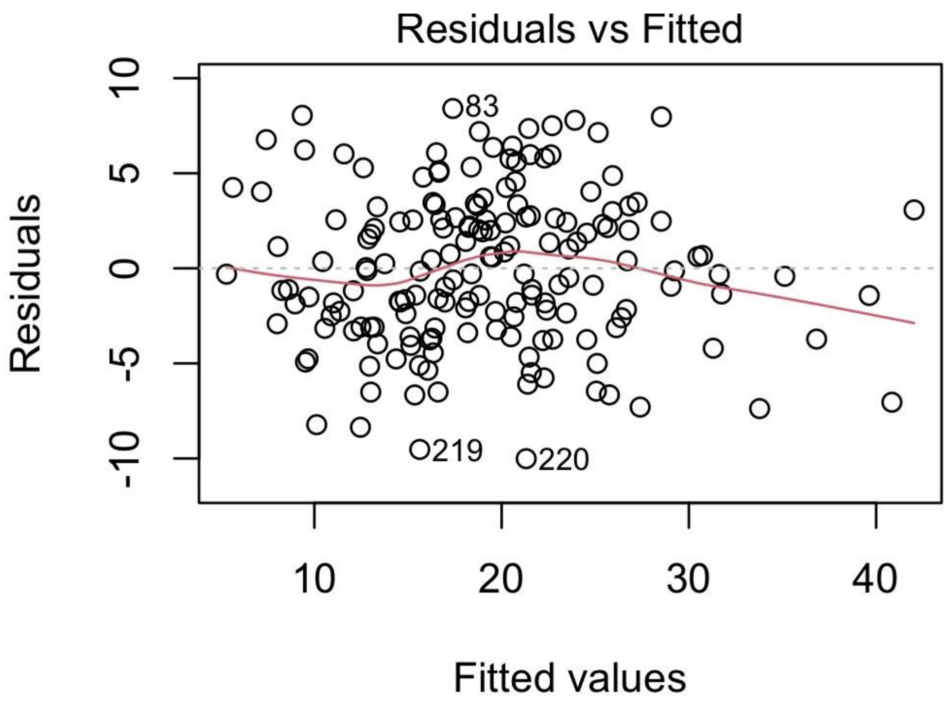
Analysis:

4. The negative coefficients for weight and wrist measurements follow a certain logic that increases in weight or wrist size may come from sources other than fat. These increases may be due to increased muscle mass or increased bone density, resulting in a decrease in body fat relative to body weight.

Model Diagnostics

- **Q-Q plot:** Although slightly skewed at both ends, the residual is approximately a normal distribution.
- The **residuals** "randomly bounce" around the 0 line: relationship is linear.
- The **residuals** form a "horizontal band" roughly around the 0 line: the variance of the error terms is equal.





Model Strengths & Weaknesses

Strength:

1. Relatively simple and easy to use.
2. Straightforward to understand and explain.
3. Satisfies the linear regression assumptions

   ——prediction results are reliable.

Weakness:

1. Data divided into training & testing sets

   —— some information may be lost when the model is built.
2. Only for normal males.