Group 15 Presentation

Bodyfat prediction project

Chufan Zhou Zhanpeng Xu Xuelan Qian

10/20/2022

Background

People are paying more and more attention to their health and bodyfat nowadays.

Accurate bodyfat measurement is inconvenient and costly.

Our group wants to build a simple, robust and accurate model to predict bodyfat based on some simple body circumference measurements

Data Cleaning

- We consider handling the individuals whose values are unusual first.
- We pay attention to the inconsistency of bodyfat vs. density and weight, height vs. adiposity(BMI) and remove those highly inconsistent.
- We change unit of height from inches to cm.

Final cleaned data:

n=244 (from n=252) with p = 14 predictors

Predictors: WEIGHT, ABDOMEN, ADIPOSITY, FOREARM,…

Data Cleaning

Individual (IDNO)	Original Obs. (bodyfat/BMI)	Reason for deletion
182	0	unusual bodyfat value
172	1.9	unusual bodyfat value
48	6.4	Inconsistency bodyfat
76	18.3	Inconsistency bodyfat
96	17.3	Inconsistency bodyfat
42	29.9	Inconsistency BMI
163	24.4	Inconsistency BMI
221	24.5	Inconsistency BMI

Building Model

- **Metric for Model Performance**

  We'll define the "best" model based on the following criteria:

  1. Simplicity of the model (number of predictors)
  2. R^2

Building Model

- **Variable selection:** Stepwise regression

For each step, we consider add one new predictor and remove one predictor in our model.

$$F = \frac{SSE_{old} - SSE_{new}}{SSE_{new}}$$

- **Candidate Models:**

Linear models searched by stepwise method

1. Y ~ ABDOMEN
2. Y ~ ABDOMEN+WEIGHT
3. Y ~ ABDOMEN+WEIGHT+WRIST
4. Y ~ ABDOMEN+WEIGHT+WRIST+FOREARM

Candidate Models

Model	R-squared
Y ~ ABDOMEN	0.6555
Y ~ ABDOMEN+WEIGHT	0.7186
Y ~ ABDOMEN+WEIGHT+WRIST	0.7253
Y ~ ABDOMEN+WEIGHT+WRIST+FOREARM	0.7339

Discussion of Candidate Models

- Y ~ ABDOMEN+WEIGHT model is **comparable** to other more complicated models; the r-squared only lower by about 2% at most.

- Model Y ~ ABDOMEN+WEIGHT has better than model Y ~ ABDOMEN+WEIGHT+WRIST+FOREARM in terms of complexity.

- Model Y ~ ABDOMEN+WEIGHT has better than model Y ~ ABDOMEN with the r-squared increase by 6%.

- The predictor ABDOMEN is significant across all models evaluated.

Final Model

Bodyfat% = - 40.27 + 0.91*Abdomen(cm) - 0.14*Weight(lb)

- Fixing abdomen, as men's weight increases by one lb, he is expected to lose 0.14 % in body fat.

- Fixing weight, as men's abdomen increases by one cm, he is expected to gain 0.91 % in body fat.

Final Model Usage

- Usain Bolt (fitnessofmens, 33 inches waist, 207 lb weight): 7.1% with 95% PI (-1.1%,15.3%)

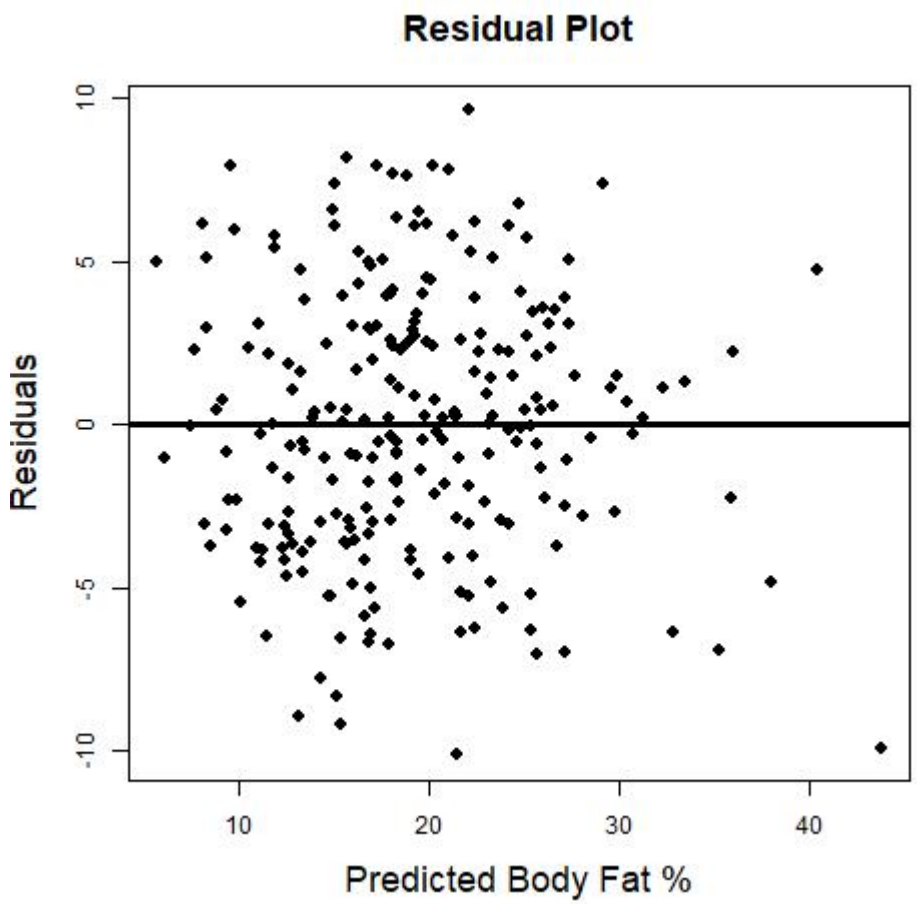
- Average American (CDC, 40.5 inch abdomen, 199.8 lb weight): 25.5% with 95% PI (17.5%,33.5%)

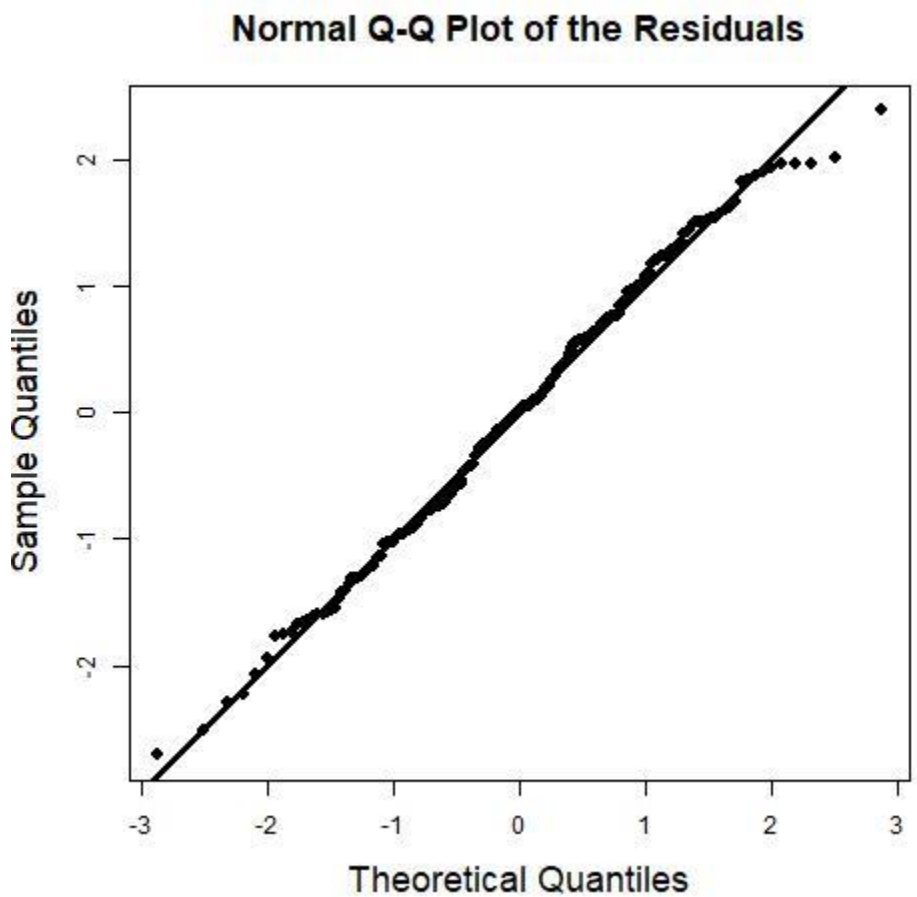
Athletes	Good	Acceptable	Overweight	Obese
5%~10%	11%~14%	15%~20%	21%-24%	25%+

Statistical Properties of Final Model

- Coef abdomen and weight all significant at 0.05 based on two-sided t-test with p-values < 10^-11.
- Coef weight is significant and negative, suggesting when fixing abdomen, weight increase mainly reflect muscle gain.
- R^2 = 0.7186, the model explains about 71.86% of the variation in body fat %.

Model Diagnostics





Strengths and Weaknesses

Final Model:

BodyFat(%) = - 40.27 + 0.91*Abdomen(cm) - 0.14*Weight(lb)

- **Strengths**
  - Very simple (abdomen in cm and weight in lb)
  - Explains 71.86% of variation in body fat
  - Linearity seems reasonable based on residual plot

- **Weaknesses**
  - Prediction is not accurate:
    - 20% of predictions within +/- 5% of true value
    - 50% of predictions within +/- 15% of true value
  - Requires units (inches or cm, lb or kg)

Thank you!