Body Fat Prediction

Chao Chang, Yezhou Li, Shuyang Chen, Ping Yu

University of Wisconsin Madison

October 10, 2019

Background and Goal

- An accurate measurement of body fat is costly and it is desirable to have easy methods of estimating body fat that are convenient.
- Come up with a simple, robust and precise model with **two** body measurements to calculate body fat.

Data Cleaning: Examine Outliers

- The 216*th* example
  - He has the largest body fat of 45.1% and his density is 0.995.
  - Delete this point.

- The 182*nd* example
  - He has a body fat of 0.
  - His body fat is -3.16 computed by the bodyfat-density formula, which is impossible.
  - Delete this point.

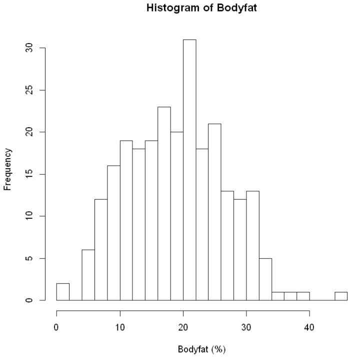

Figure: Histogram of body fat

Data Cleaning: Examine Outliers

- The 42*nd* example has the shortest height of 29 inches but his other measurements are normal
- Impute his height using ADIPOSITY (BMI) formula. $\Rightarrow$ 69.4 inches.
- Keep this point.

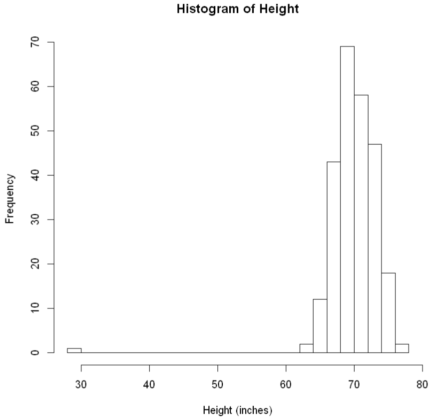

Figure: Histogram of height

Data Cleaning: Examine Outliers

- The 39*th* man is the fattest person with the largest weight, abdomen, chest, hip, thigh, knee, biceps, wrist and adiposity.
- Keep this point.

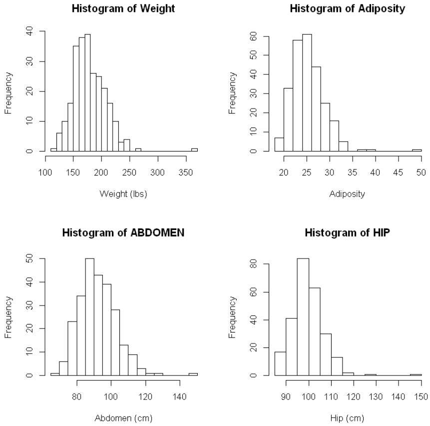

Figure: Histogram of weight, adiposity, abdomen and hip

Data Cleaning: Examine Outliers

We compute the **ADIPOSITY (BMI)** using weight and height to see whether there are some wrong records in this variable:

ID	Difference
163	3.007
221	2.822
156	0.306
235	0.242
116	0.236
86	0.230

Table: Difference between records and computed value of Adiposity

- The 163*rd* and 221*st* examples' difference are much larger than others.
- Delete these two points.

Data Cleaning: Final Data Set

- Delete $163rd, 182nd, 216th, 221st$ examples.
- The final data set has 248 examples and 14 variables.

Models

- Linear model:

$$Bodyfat_i = \alpha_0 + \alpha_1 ABDOMEN_i + \alpha_2 WRIST_i + \epsilon_i, \epsilon_i \sim N(0, \sigma_1^2)$$

- Fraction model:

$$Bodyfat_i = \frac{\beta_1 ABDOMEN_i + \beta_2}{WEIGHT_i} + \gamma_i, \gamma_i \sim N(0, \sigma_2^2)$$

Linear Model: Variable Selection

- Stepwise variable selection
- Criterion: BIC
  - Models selected by AIC have too many variables.
  - Using BIC, both forward and backward selection give us a model with four variables:

$$Bodyfat \sim Weight + Abdomen + Forearm + Wrist$$

Linear Model: Collinearity Problem

The VIF values of four predictors:

Predictor	Weight	Abdomen	Forearm	Wrist
VIF	7.70	5.51	1.70	2.12

Table: VIF of four predictors

- **Weight** has some collinearity problem.
- Drop weight to see whether the model performance changes a lot.

Linear Model: Final model

Divide the full data set into training (70%) and test set (30%).

Models	RMSE	R^2
Four variables	4.17	0.71
Abdomen + Forearm + Wrist	4.33	0.69
Forearm + Wrist	**7.08**	**0.17**
Abdomen + Wrist	4.33	0.69
Abdomen + Forearm	4.34	0.68
Weight + Abdomen	4.27	0.70

Table: Model performance

- The best linear model (using the full data set) is:

  $$\text{Bodyfat} = 0.682 \times \text{Abeomen} - 2.022 \times \text{Wrist} - 7.256, \; R^2 = 0.6832$$

Fraction Model: Logics

- There is a linear relationship between body fat and $\frac{1}{Density}$.

$$Bodyfat = \frac{495}{Density} - 450$$

- By physics theorem:

$$\frac{1}{Density} = \frac{Volume}{Weight}$$

- We already have accurate measurement of weight and we try to build a linear model to estimate volume.
- Restricted by the number of predictors we can use (2), the model we try to construct is:

$$Bodyfat_i \sim \frac{\beta_1 X_i + \beta_2}{Weight_i} + \beta_3$$

Fraction Model: Variable Selection

Use the 10-fold cross validation, we iterate through all 14 variables.

- No mater we sort by RMSE or the R^2, **Abdomen** stands out among all of the variables.

Varible	RMSE	R^2
ABDOMEN	**4.06**	**0.72**
ADIPOSITY	5.10	0.55
CHEST	5.41	0.51
AGE	5.51	0.47
WRIST	5.73	0.43
HEIGHT	5.75	0.51
HIP	5.76	0.43
THIGH	5.84	0.40
NECK	5.87	0.41
ANKLE	5.88	0.41
KNEE	5.90	0.40
FOREARM	5.92	0.39
BICEPS	5.92	0.39

Table: Model performance

Fraction Model

After we fit the model using the full data, the result is the following with $R^2 = 0.7089$:

	Estimate	Std. Error	p value
Intercept	-2.515	3.559	0.48
invWeight	-10736.297	438.877	$< 2e-16^{***}$
ABD-WEI	158.986	9.566	$< 2e-16^{***}$

The intercept is not significant so we try to drop the intercept. The model without the intercept has $R^2 = 0.9605$.

	Estimate	Std. Error	p value
invWeight	-10632.653	413.229	$< 2e-16^{***}$
ABD-WEI	153.050	4.573	$< 2e-16^{***}$

Fraction Model: Final Model

The best fraction model is:

$$Bodyfat_i = \frac{153.05\text{Abdomen}_i - 10632.653}{\text{Weight}_i}, R^2 = 0.9605$$

Final Model

	RMSE		R^2	
Model	10-fold	full data	10-fold	full data
Linear model	4.20	4.207	0.702	0.6832
Fraction model	4.03	4.036	0.723	0.9605

Table: Comparison between linear and fraction model

The final model is:

$$Bodyfat_i = \frac{153.05\text{Abdomen}_i - 10632.653}{\text{Weight}_i}$$

Model Diagnose

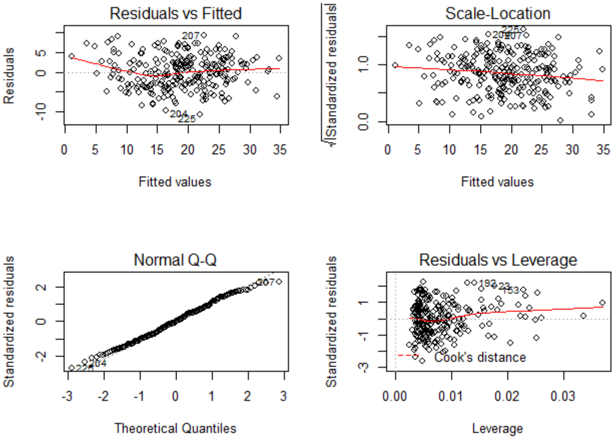

Figure: Diagnose plots of fraction model

Our Rule of Thumb

Rule of thumb

$$\text{Bodyfat}_i = \frac{153 \times \text{Abdomen}_i - 1.06 \times 10^4}{\text{Weight}_i}$$

Interpretation: 153 times the ratio between abdomen and weight, and then minus 1.06×10^4 times the inverse of weight.

Shinny Application

Our shinny application URL is:
https://clarefrost.shinyapps.io/shiny_628_group4/

Thank You