Men's Bodyfat Percentage Calculator

Yunhui Qi, Sam Waterbury, Junxia Zhu

University of Wisconsin – Madison

February 7, 2019

Overview

Rule of Thumb and Example

Rule of Thumb

$$BODYFAT = 0.91 \cdot ABDOMEN - 0.14 \cdot WEIGHT - 40.27$$

"90% abdomen minus 10% weight, don't forget to minus 40%, get your body fat!".

Example Usage

for an ordinary man with abdomen 84cm and weight 173lbs, his predicted body fat % percentage would be around 12.05%.

Data Description

IDNO	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
1	12.6	1.0708	23	154.25	67.75	23.7	36.2	93.1	85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
2	6.9	1.0853	22	173.25	72.25	23.4	38.5	93.6	83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
3	24.6	1.0414	22	154.00	66.25	24.7	34.0	95.8	87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6
4	10.9	1.0751	26	184.75	72.25	24.9	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2
5	27.8	1.0340	24	184.25	71.25	25.6	34.4	97.3	100.0	101.9	63.2	42.2	24.0	32.2	27.7	17.7
6	20.6	1.0502	24	210.25	74.75	26.5	39.0	104.5	94.4	107.8	66.0	42.0	25.6	35.7	30.6	18.8

- 252 observations of 16 features
- *DENSITY* is measured accurately from underwater weighing and *BODYFAT* is obtained by Siri's equation.

Data Cleaning

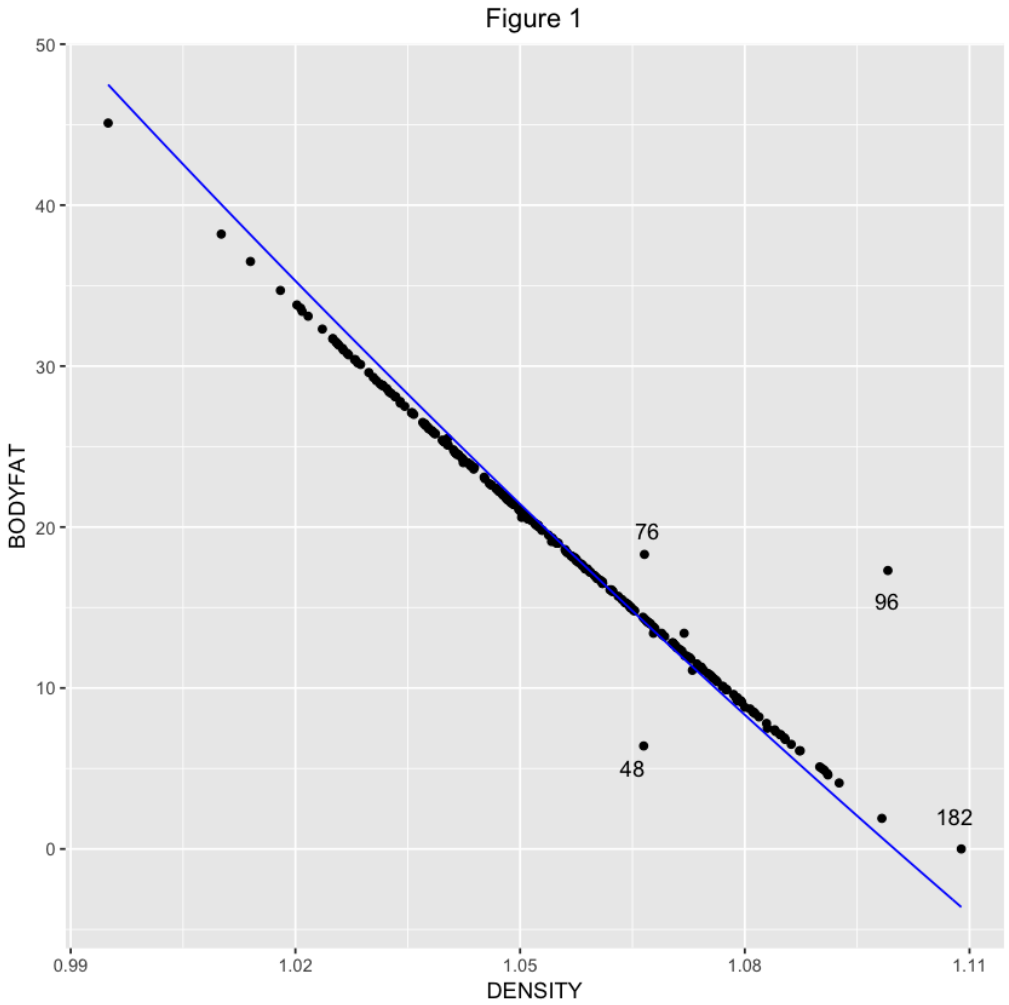


Figure: linear diagnostic plot

IDNO	BODYFAT	DENSITY	AGE
48	6.4	1.0665	39
76	18.3	1.0666	61
96	17.3	1.0991	53
182	0.0	1.1089	40

Figure: abnormal points

BFP calculated by Density

- 48^{th}: 14.1%
- 76^{th}: 14.1%
- 96^{th}: 0.37%

Data Cleaning

IDNO	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY
182	0	1.1089	40	118.5	68	18.1

Figure: 182^{nd} point

- 0% of bodyfat is impossible, and the BODYFAT calculated by DENSITY is negative.
- ADIPOSITY measures BMI which relies on weight and height.Check if this is correct by $BMI = 703 * Weight(lbs)/Height^2(in^2)$
- $AdultBodyFat\% = 1.51 \cdot BMI - 0.70 \cdot AGE - 3.6 \cdot SEX + 1.4$ Impute BFP of 182^{nd} point with value 14.7%

Data Cleaning

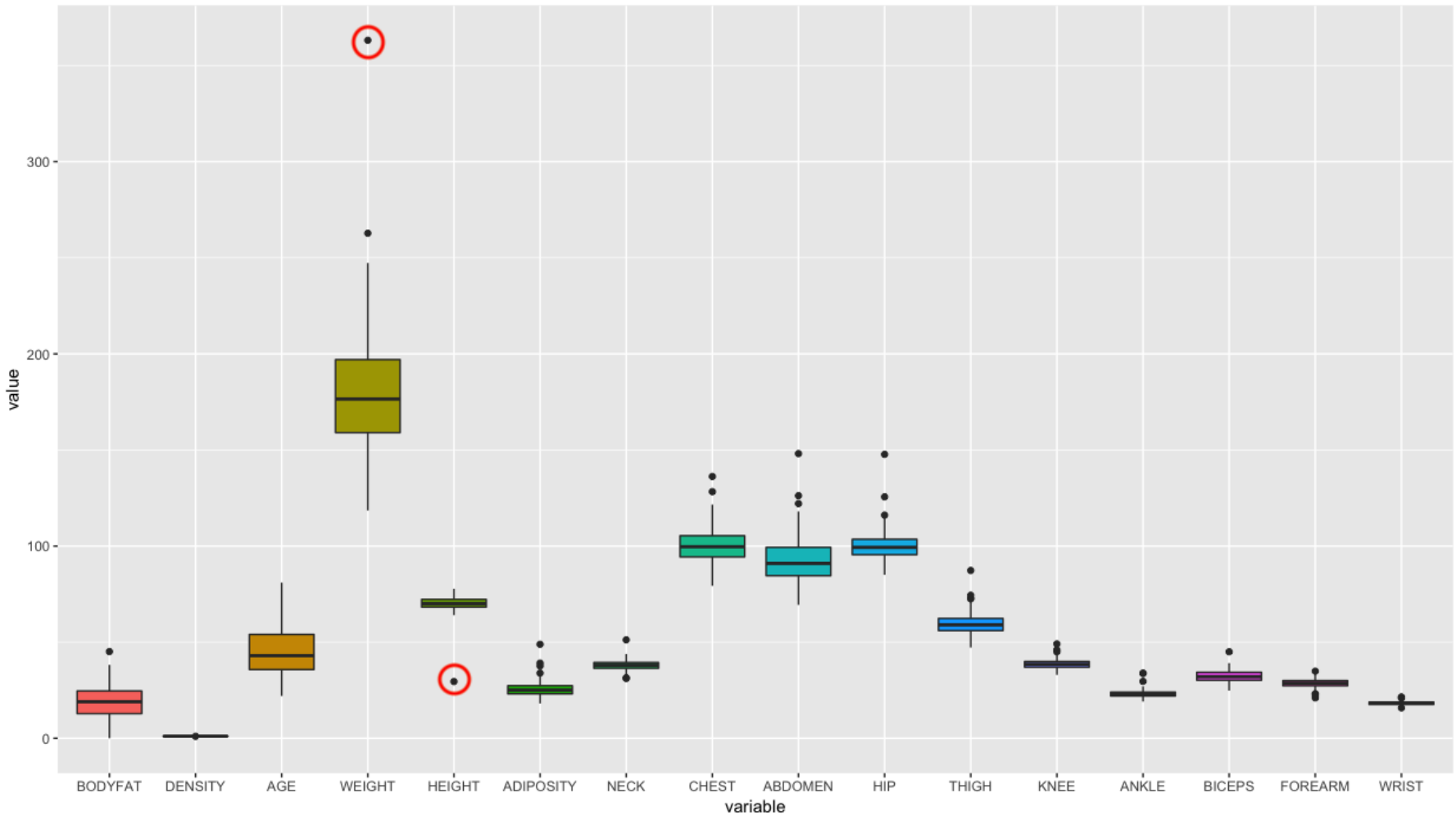


Figure: Boxplot on all variables

Data Cleaning

IDNO	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN
39	33.8	1.0202	46	363.15	72.25	48.9	51.2	136.2	148.1
42	31.7	1.0250	44	205.00	29.50	29.9	36.6	106.0	104.3

Figure: abnormal points

- 39^{th}: Weight is extremely high.
- 42^{th}: Height is 29.50 inches. Apply $BMI = 703 * Weight(lbs)/Height^2(in^2)$ and got 69.43 inches.
- We also detect some other points, decide not to do any changes.

Variable Selection

- **Stepwise Selection**
  - Bidirectional stepwise regression using both AIC and BIC
  - AIC resulted in models with too many features
  - BIC identified 4 important variables: *ABDOMEN*, *WEIGHT*, *WRIST*, and *FOREARM*
- **Lasso Regression**
  - Chose λ-value which selected a model with four variables
  - It selected *AGE*, *HEIGHT*, *ABDOMEN*, and *WRIST*
- Thus, we narrowed our search to six variables: *ABDOMEN*, *WRIST*, *WEIGHT*, *FOREARM*, *AGE*, and *HEIGHT*

Variable Selection

- Using these six variables, we performed **Best Subset Selection**
- This found the best variables for a model of size $p = 1, 2, 3, 4$:

p	$\sqrt{\text{MSE}}$	Features
1	4.54	*ABDOMEN*
2	4.14	*ABDOMEN*, *WEIGHT*
3	4.08	*ABDOMEN*, *WEIGHT*, *WRIST*
4	4.04	*ABDOMEN*, *WEIGHT*, *WRIST*, *FOREARM*

- We chose to use $p = 2$ since it provides the best balance between simplicity and accuracy
- We also performed BSS using second-order interactions (multiplication and ratios), but this did not improve the accuracy of an estimator with 2 variables

Model Fitting

- **SLR model:**

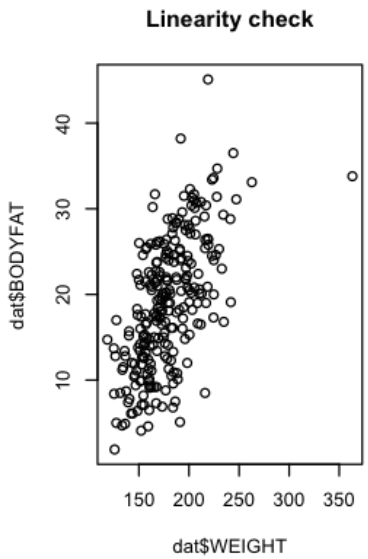
$$BODYFAT\% = 0.91 \cdot ABDOMEN(lbs) - 0.14 \cdot WEIGHT(cm) - 40.27\%$$

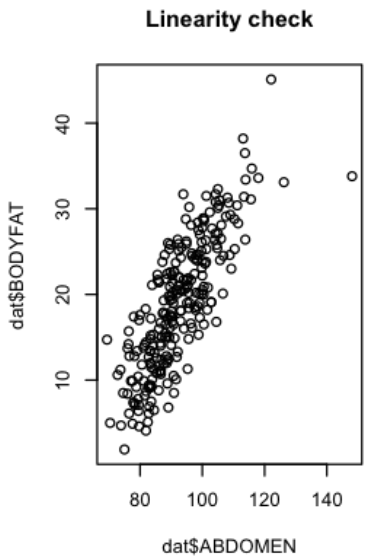
- **Interpretation of SLR results:**
  - Coefficients:
    - Negative coefficient of weight: gain of weight = fat or muscle ?
  - R^2: 70.98% of the variation in body fat,compared with the full model which has adjusted R^2 equal to 72.49%, our model doesn't have much efficiency loss but less predictors.

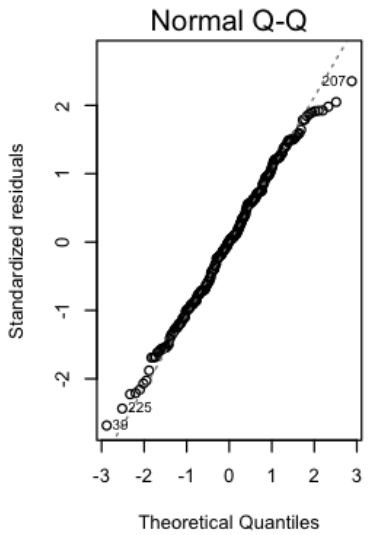
Model Diagnostic

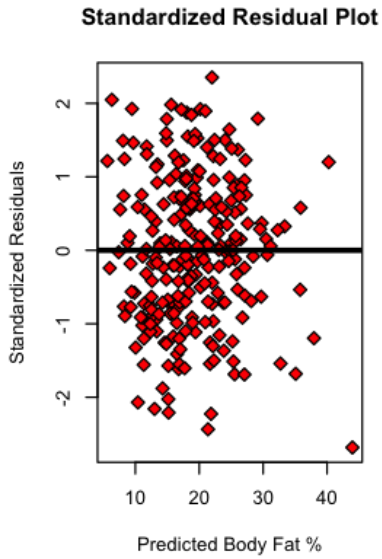
1. Additivity: reasonable from the interpretation of coefficients
2. Constant effects: have no reason to deny.
3. Fixed X: have no reason to deny.
4. Multicolinearity: vif value less than 10.
5. Linearity: scatter plot, fitted vs residuals plot
6. Error terms: Q-Q plot,fitted vs residuals plot, standardized residuals plot.
7. Outliers: cook's distance and pii values, leverage plot.

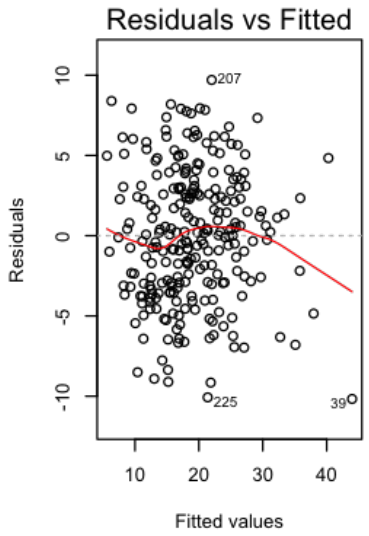
Model Diagnostic

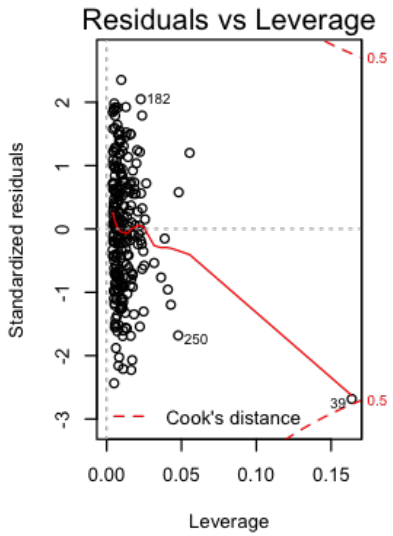












Robustness Tests

- Trade off between precision and robustness.
- Bootstrap sample of original data set.

	with	without	bootstrap
(Intercept)	-40.27	-41.77	-40.27
ABDOMEN	0.91	0.90	0.91
WEIGHT	-0.14	-0.12	-0.14

Strengths and Weaknesses

- **Strengths:**
  - Adopting background information to impute strange points.
  - Considering interaction and ratio terms.
  - Evaluation of robustness and precision.
- **Weakness:**
  - P hacking problem.
  - Negative coefficient of weight gives some constrain to our shiny app.

The End