

# Bodyfat prediction

Y. Tang, H. Xia, Y. Zheng

University of Wisconsin-Madison

February 4, 2019

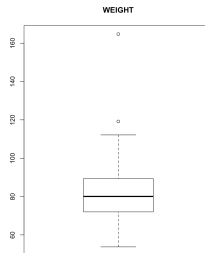
# Introduction

- dataset:  $Y$ =body fat Percentage;  $X$ =14 physical measurements
- model: a two-term nonlinear model

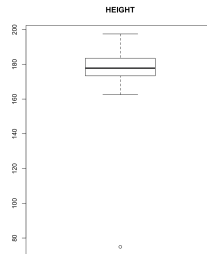
# Data Cleaning

- Summary of variables

BODYFAT		AGE	
Min.	: 0.00	Min.	:22.00
1st Qu.	:12.80	1st Qu.	:35.75
Median	:19.00	Median	:43.00
Mean	:18.94	Mean	:44.88
3rd Qu.	:24.60	3rd Qu.	:54.00
Max.	:45.10	Max.	:81.00



WEIGHT		HEIGHT	
Min.	: 53.75	Min.	: 74.93
1st Qu.	: 72.12	1st Qu.	:173.35
Median	: 80.06	Median	:177.80
Mean	: 81.16	Mean	:178.18
3rd Qu.	: 89.36	3rd Qu.	:183.51
Max.	:164.72	Max.	:197.49



- four weird records

---

39	weight=164.72 lbs
42	height=74.93 inch
79	age=81 years old
182	bodyfat=0

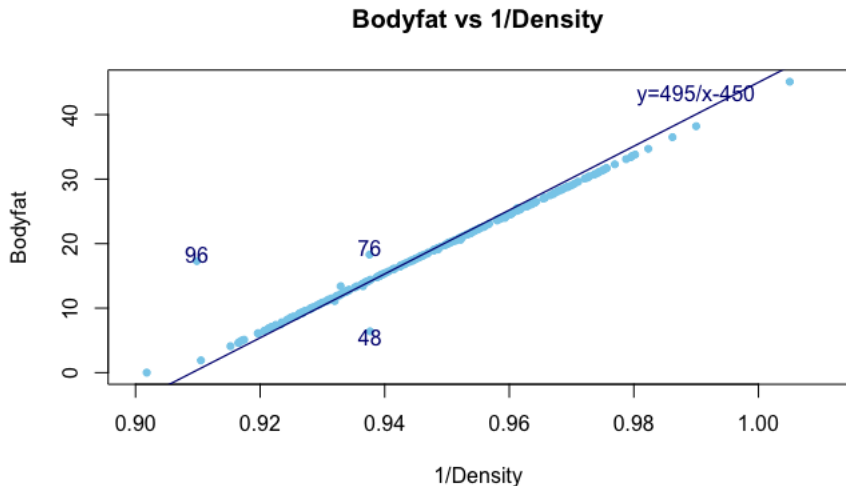
- Consistency of BMI versus HEIGHT and WEIGHT

$$BMI = \frac{WEIGHT}{(\frac{HEIGHT}{100})^2}$$

We change the height of 42th but retain 39th and 79th because it seems that the 79th is a normal thin old man, the 39th is a very heavy man which follows the bmi equation.

# Data Cleaning

- Consistency between DENSITY and BODYFAT



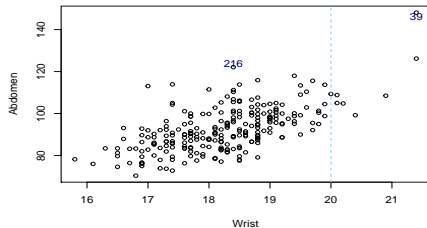
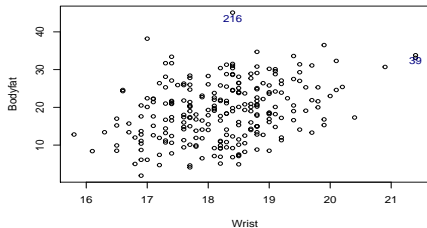
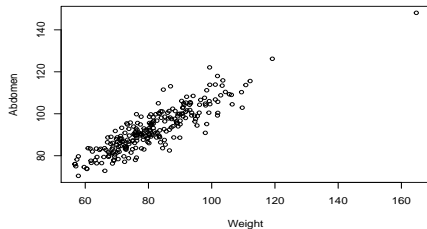
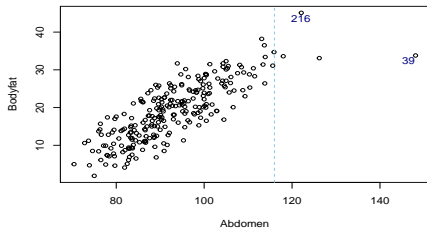
# Variable Selection

Forward stepwise variables selection based on AIC

Num. of Variables	Variables Selected	AIC
1	Abdomen	757.29
2	Abdomen, Weight	712.53
3	Abdomen, Weight, Wrist	706.45
4	Abdomen, Weight, Wrist, Forearm	702.18

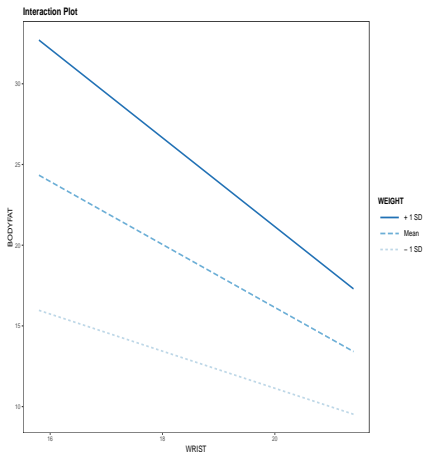
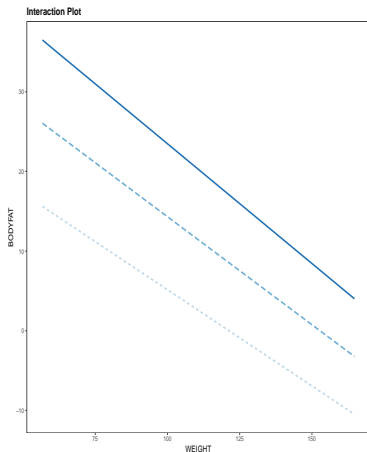
# Bodyfat, Abdomen, Weight, Wrist Relationships

Variables Relationships





# Interaction Effect



# Model Comparison

Model	Adj.R-squared	MSE
m2_ridge	0.7127	16.69
m3	0.7185	16.16
m5	0.7258	15.80

# Model Comparison

- Final Model

$$\text{Bodyfat}(\%) = -45.3249 + 0.9092 \times \text{Abdomen}(\text{cm}) \\ - 0.0133 \times \text{Weight}(\text{kg}) \times \text{Wrist}(\text{cm})$$

	Estimate	Std.Error	t.value	P.value
Intercept	-45.3249	2.573	-17.62	2e-16
<i>ABDOMEN</i>	0.9092	0.047	19.42	2e-16
<i>WEIGHT : WRIST</i>	-0.0133	0.002	-8.22	1.14e-14

# Model Diagnostic

- Normality Assumption

$H_0$ : the residual follows normal distribution.

- Homoscedasticity Assumption

$H_0$ : homoscedasticity vs  $H_1$ : variance residuals vary with the level of fitted values

---

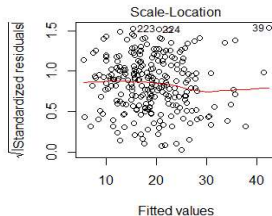
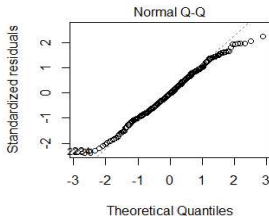
Shapiro-Wilk normality test

$W = 0.99056$   
 $p\text{-value} = 0.1039$

Non-constant Variance Score Test

$\text{Chisquare} = 0.0001924752, \text{Df} = 1$   
 $p = 0.98893$

---



- Robustness

`rlm(formula = BODYFAT ~ ABDOMEN + WEIGHT : WRIST)`

	coefficients from m5	coefficients from rlm5
Intercept	-45.325	-46.652
<i>ABDOMEN</i>	0.909	0.922
<i>WEIGHT : WRIST</i>	-0.013	-0.013

The coefficients of the robust model5 are very close to those of the previous model5, which means model5 is robust to some extent.

$$\begin{aligned} \text{Bodyfat}(\%) = & -45.3249 + 0.9092 \times \text{Abdomen}(\text{cm}) \\ & -0.0133 \times \text{Weight}(\text{kg}) \times \text{Wrist}(\text{cm}) \end{aligned}$$

$$\text{Bodyfat}(\%) = -45 + 0.91 \times \text{Abdomen}(\text{cm}) - 0.013 \times \text{Weight}(\text{kg}) \times \text{Wrist}(\text{cm})$$

- Explain the practical meaning of this model:

For a 75 kg man whose abdomen is 85 cm, wrist is 18 cm, his predicted bodyfat is around 14.97%. There is a 95% probability that his bodyfat is between 14.35% and 15.59%. Second model is a simpler one to calculate bodyfat, it predicts this person has 14.8% bodyfat.

*Thank you!*