Body Fat Prediction

JIA LIU, YIQIAO ZHANG

Part 1 Introduction

- Dataset:

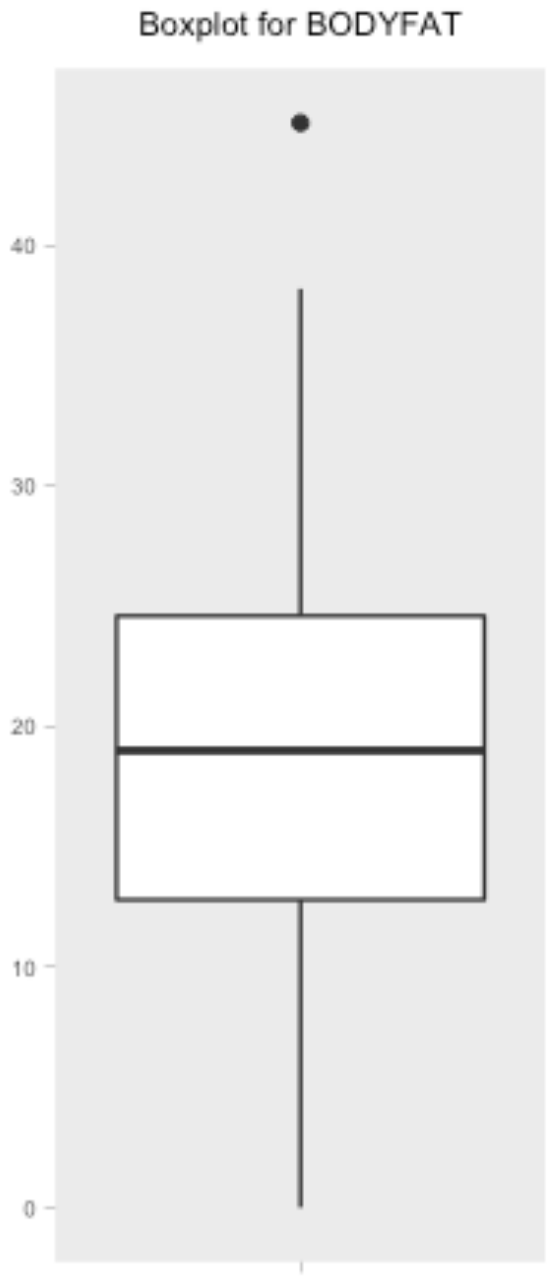
  Percentage of body fat + other 14 physical measurements of 252 men.

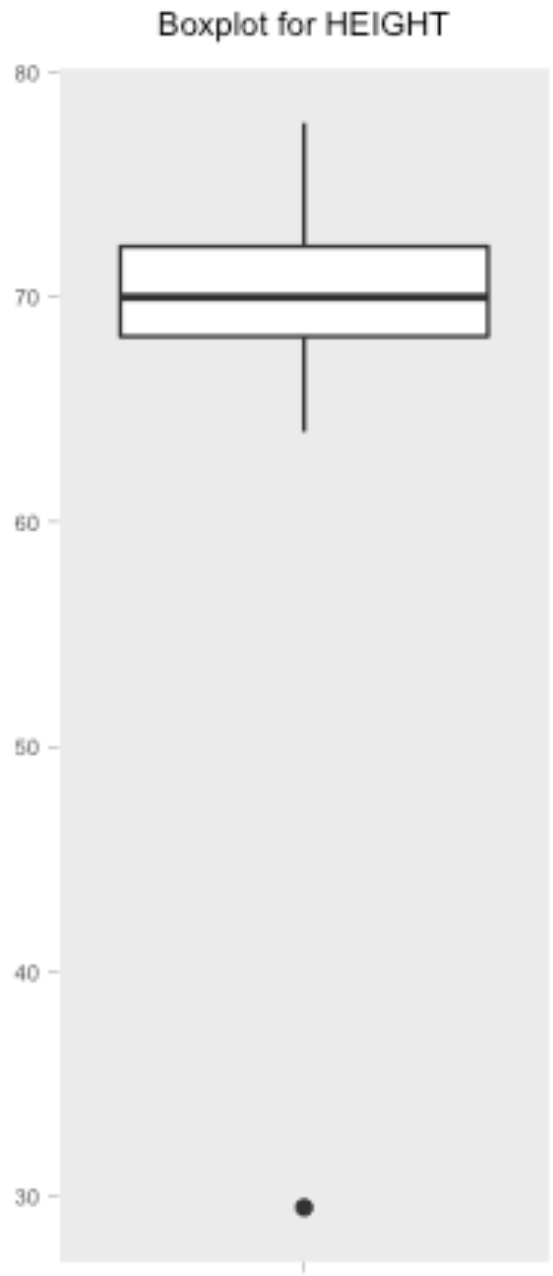
- Goal:

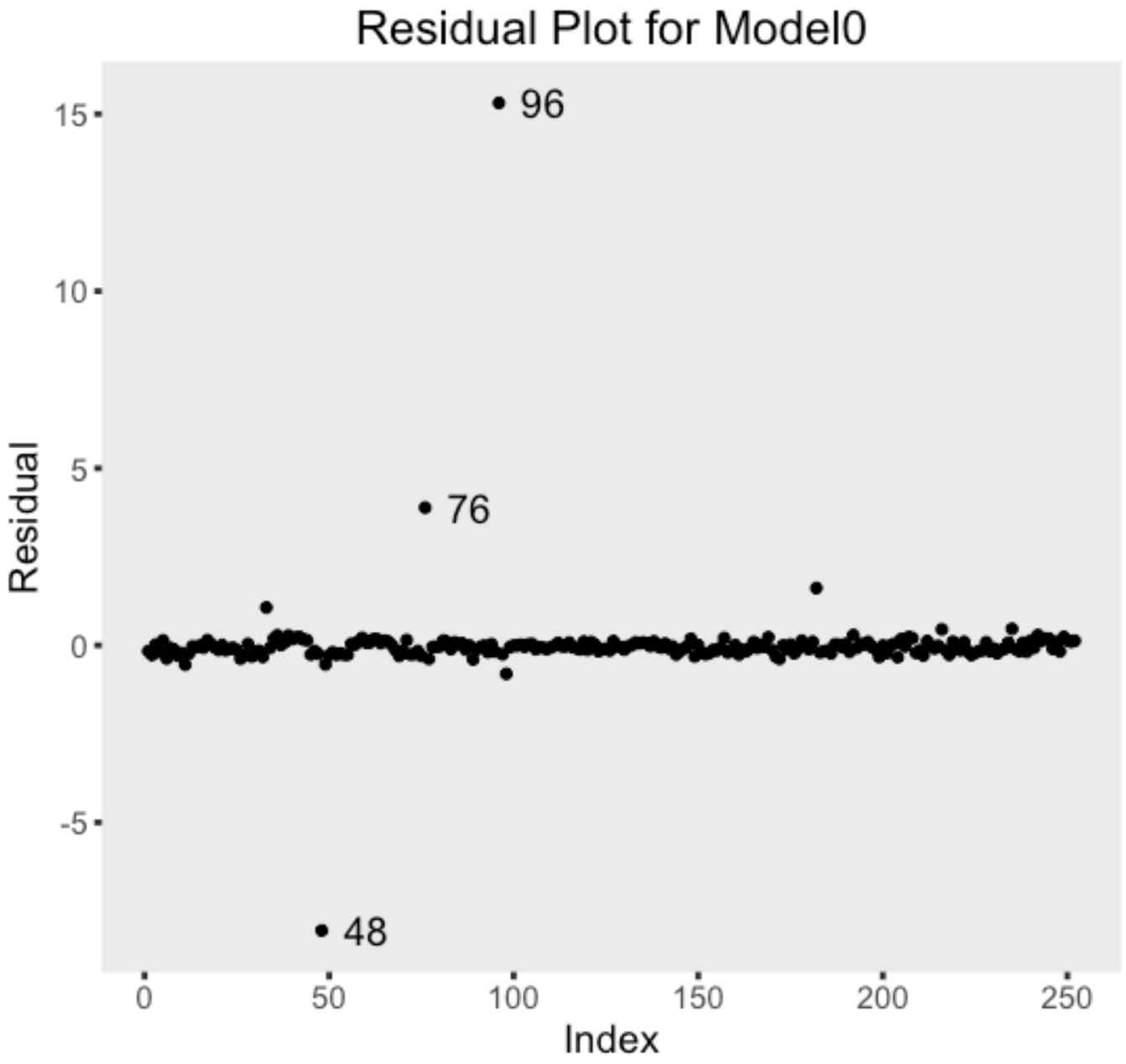
  Propose a linear model to predict body fat.

Part 2:
Data Cleaning

Boxplot of raw data:







Data Cleaning

Consistence of BODYFAT

versus DENSITY:

$$Bodyfat = \beta_0 + \beta_1 \frac{1}{Density}$$

Data Cleaning

Now let's look into these three points.

	Recorded Body Fat	Estimated Body Fat by Formula
Record No.95	9.6	0.4

This is too small, impossible!

Data Cleaning

	Recorded Body Fat	Estimated Body Fat by Formula
Record No.75	18.3	14.1

Compare with a correctly recorded point No. 24:

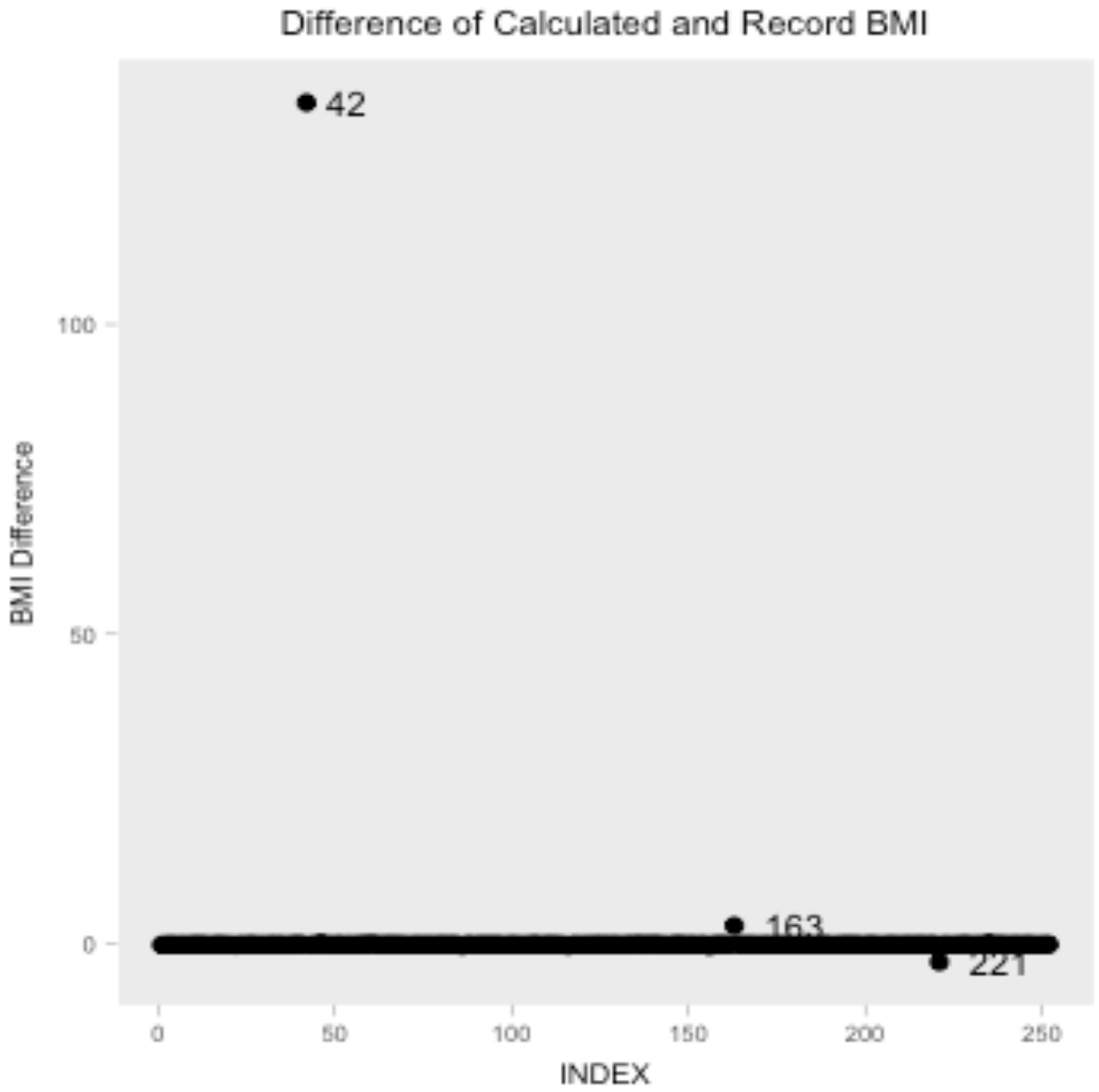
The recorded Body Fat is more reasonable, stick to that one.

Data Cleaning

	Recorded BodyFat	Estimated BodyFat by Formula
Record No.47	6.4	14.1

Compare with a correctly recorded point No. 24:

The estimated Body Fat is more reasonable, change it.



Data Cleaning

Consistence of BMI versus HEIGHT and WEIGHT:

$$BMI = \frac{weight(kg)}{height(m^2)}$$

These two points are excluded.

Part 3: Variable Selection

Method	Selected Variables
BIC Backward	WEIGHT, ABDOMEN, FOREARM, WRIST
BIC Forward & Both	ABDOMEN, WEIGHT
AIC Backward	10 variables

Part 3: Variable Selection

Applied ANOVA on selected models vs full model

“BIC backward” model with four variables is good enough:

RSS	Df	F statistics	P value
3608	NA	NA	NA
3855	-10	1.59	0.1

Part 3: Variable Selection

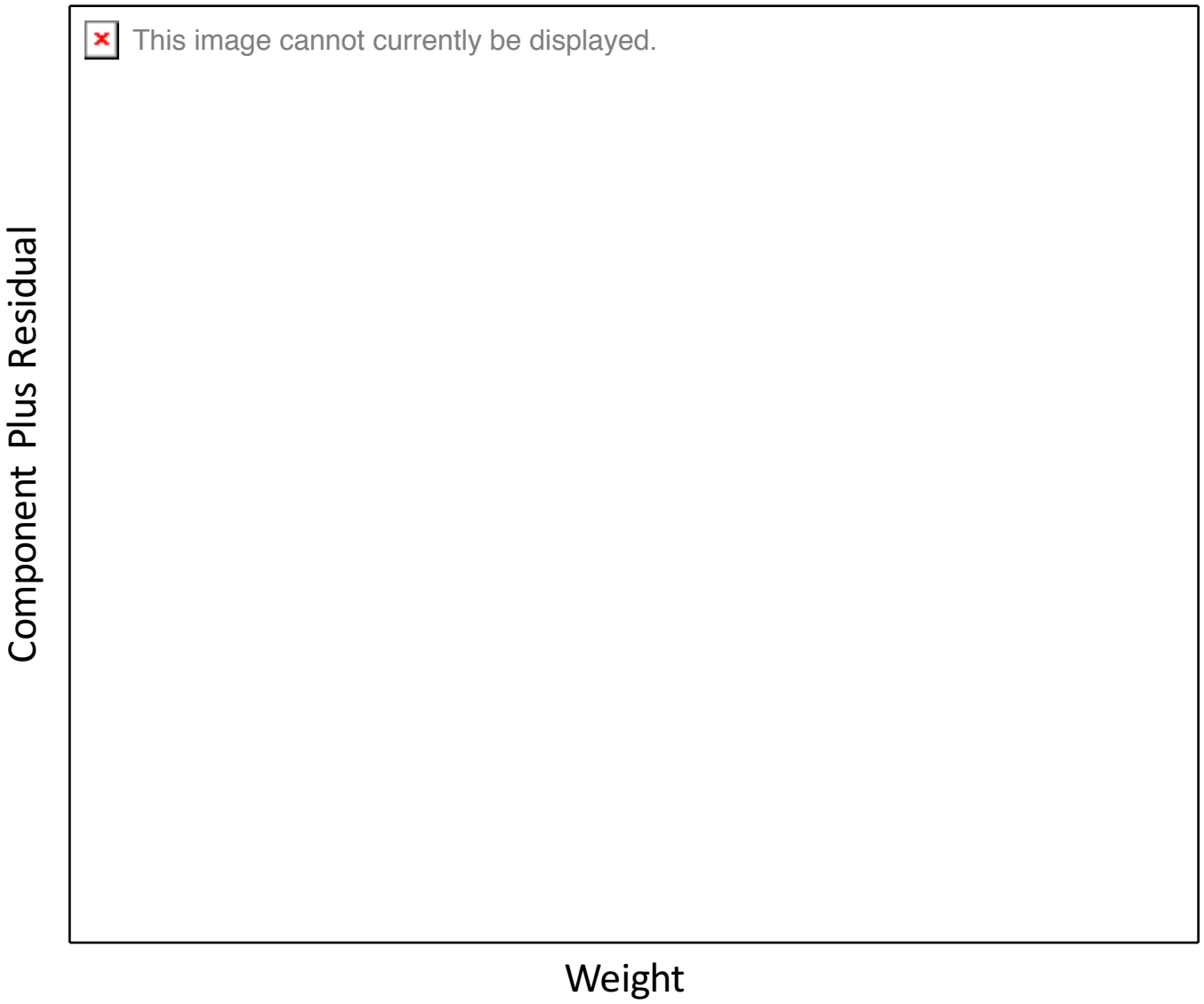
Two reserved modelS:

Model 1:

$$Bodyfat \sim WEIGHT + ABDOMEN + FOREARM + WRIST$$

Model 2:

$$Bodyfat \sim WEIGHT + ABDOMEN$$



Part 4: Model Diagnostic

For model 1:

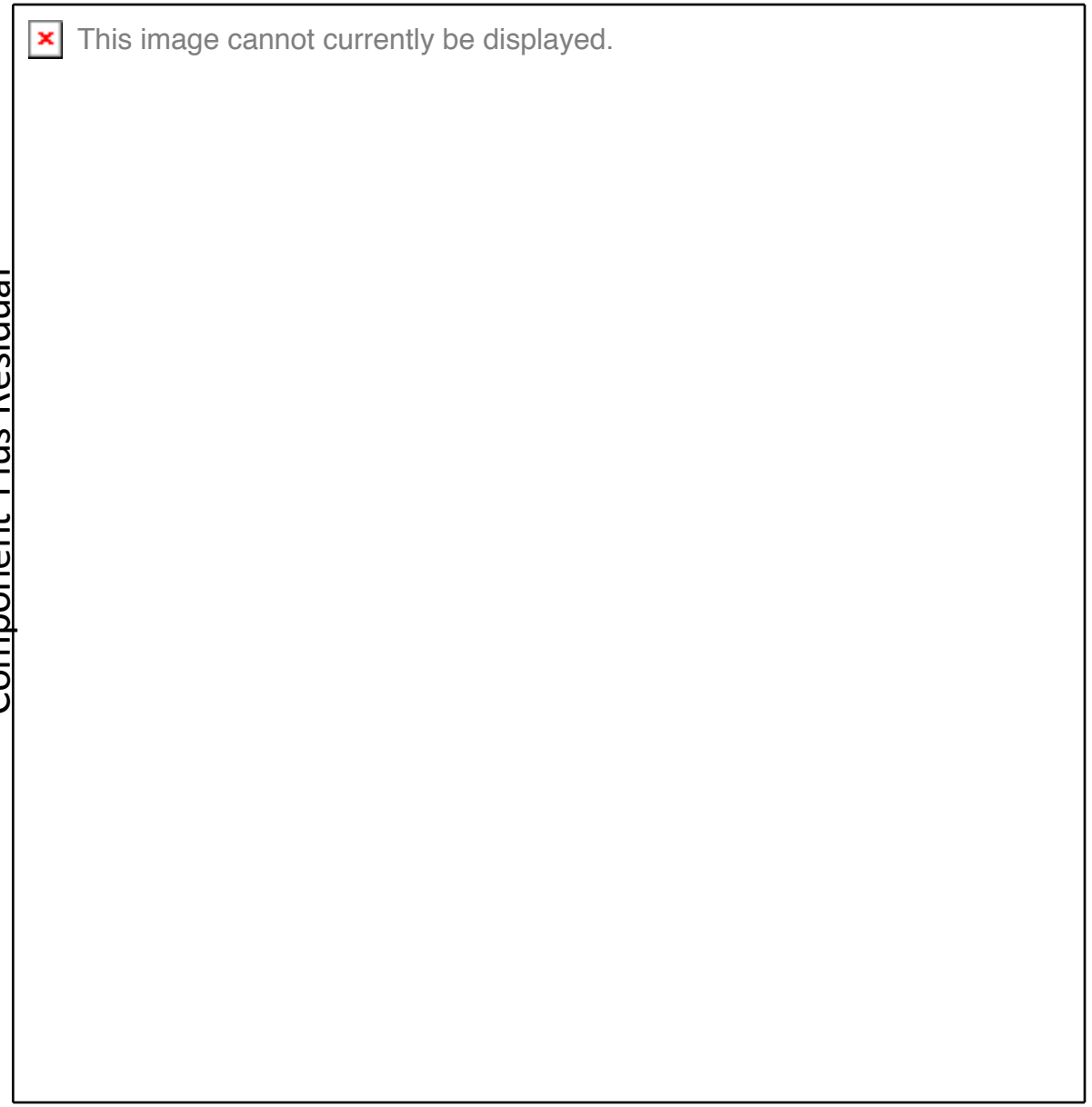
We found that linearity assumption is not satisfied:

Model Diagnostic

For model 1:

After transformation, we got:

$$Bodyfat(\%) = 30 - 0.5WEIGHT(kg) + 0.9ABDOMEN(cm) + 0.2FOREARM(cm) - 1.3WRIST(cm) - 766.5WEIGHT^{\{-0.7\}}(kg)$$



Model Diagnostic

For model 2:

Again, linearity assumption is not satisfied.

Model Diagnostic

For model 2:

After transformation, we got:

$$Bodyfat(\%) = 17.42 - 0.58WEIGHT(kg) + 0.91ABDOMEN(cm) - 775.34WEIGHT^{\{-0.7\}}$$

Part 5: Model Comparison and Outlier Detection

The adjusted R-square of four models we got:

2 variables	2 variables with transformation	4 variables	4 variables with transformation
0.7398	0.729	0.7294	0.7168

They are quite similar, so we choose the simplest one!

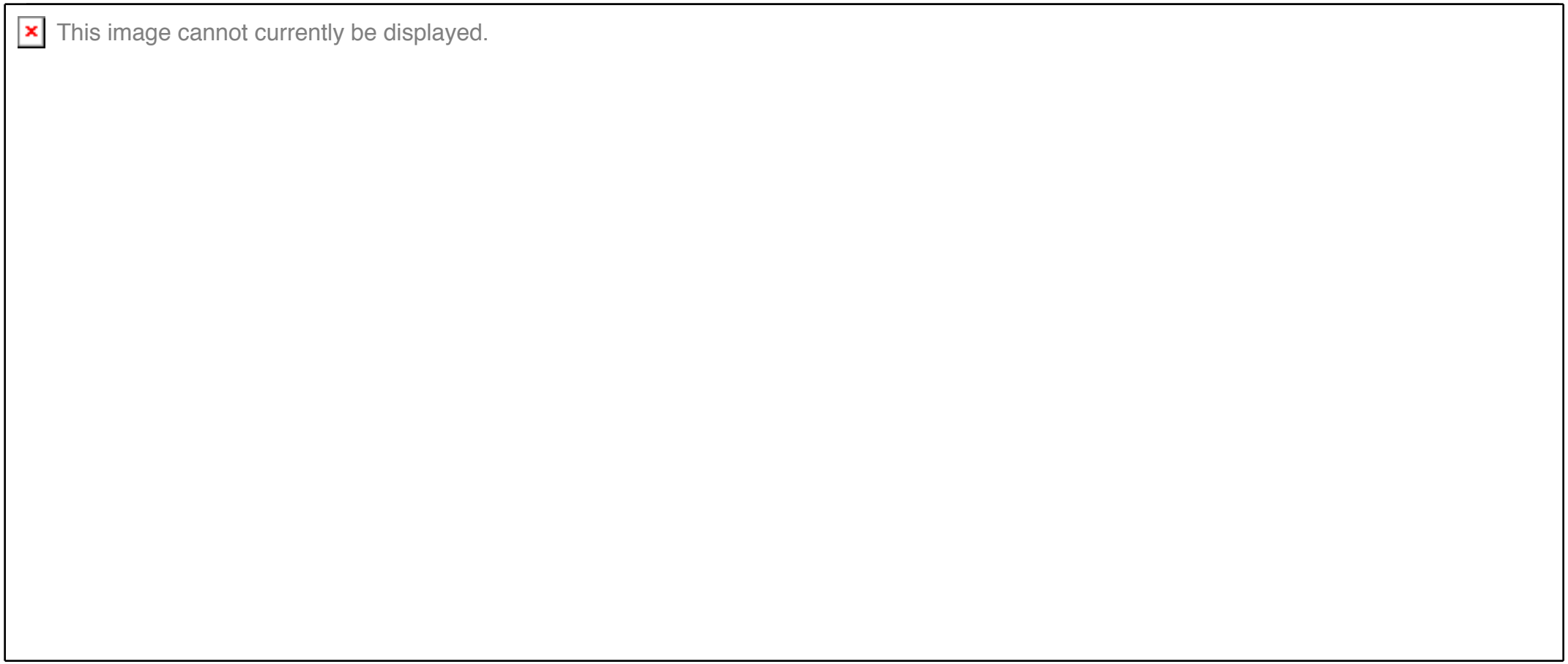
Part 5: Model Comparison & Outlier Detection

Our final model:

$$\boldsymbol{Bodyfat}(\%) = -\mathbf{40.8} - \mathbf{0.31}\boldsymbol{WEIGHT}(\boldsymbol{kg}) + \mathbf{0.92}\boldsymbol{ABDOMEN}(\boldsymbol{cm})$$

	P(>|t|)
Intercept	3.75e-43
Weight	2.90e-12
Abdomen	2.59e-45

Part 5: Model Comparison and Outlier Detection



Model Comparison and Outlier Detection

Test Results for Model 2 :

Test	Null hypothesis	P-value
Normality Test	residual is normally distributed	0.2689
Homoscedasticity Test	error's variance is constant	0.8069476

VIF:

	WEIGHT	ABDOMEN
VIF	4.68	4.68

Part 6: Rule of Thumb

"Multiply your weight (kg) by 0.3, add your abdomen (cm) and minus 50"

$$BodyFat(\%) = -50 + 0.3WEIGHT(kg) + ABDOMEN(cm)$$

Example:

WEIGHT (kg)	ABDOMEN (cm)
80	90

BodyFat result →

Precisely Estimate	95% Confidence Interval	Rule of Thumb
17%	[16.5%, 17.5%]	16%

Part 7: Strength and Weakness

Strength:

1. Constant Effects
2. Consistency of Unit
3. Intuitively Reasonable
4. Valid Model
5. Simplicity

Weakness:

1. Linearity Assumption may not satisfied
2. Model Suitable for Men only

Further work

Find an nonlinear model if exists.