# 001-Motivating Example

Body Fat Data

May 25, 2015

Abstract

Identifying overweight populations is an important first step in fighting the obesity epidemic. However, accurate measure of body fat are costly and inconvenient. Therefore we are interested in determining predictors of body fat which require only a scale and a measuring tape. We analyze a dataset which contains percentage of body fat, age, weight, height and ten body circumference measurements for 251 men. We model the data using multiple linear regression and perform various model selection techniques.

1   EDA

Simple Scatterplot Matrix of Fat data
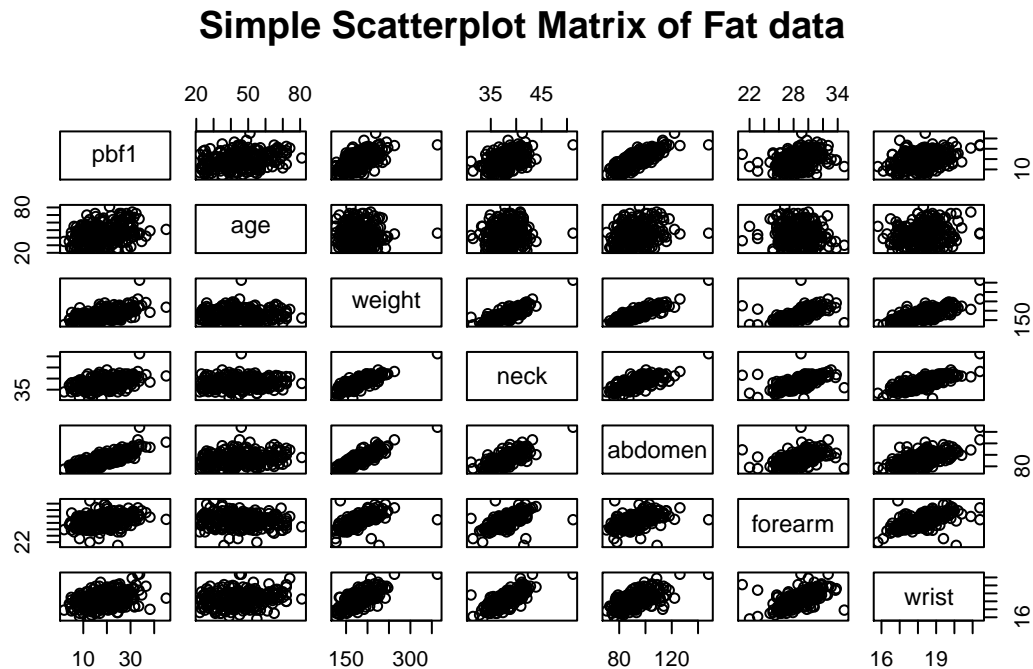
We will fit a model of the form

$$pbf1_i = \beta_0 + \beta_1\text{age}_i + \beta_2\text{weight}_i + \beta_3\text{height}_i + \beta_4\text{neck}_i$$
$$+ \beta_5\text{chest} + \beta_6\text{abdomen}_i + \beta_7\text{hip}_i + \beta_8\text{thigh}_i + \beta_9\text{knee}_i$$
$$+ \beta_{10}\text{ankle}_i + \beta_{11}\text{bicep}_i + \beta_{12}\text{forearm}_i + \beta_{13}\text{wrist}_i, \quad (1)$$

2   Results

The parameter estimates of Model (1) and their standard errors are shown in Table 1
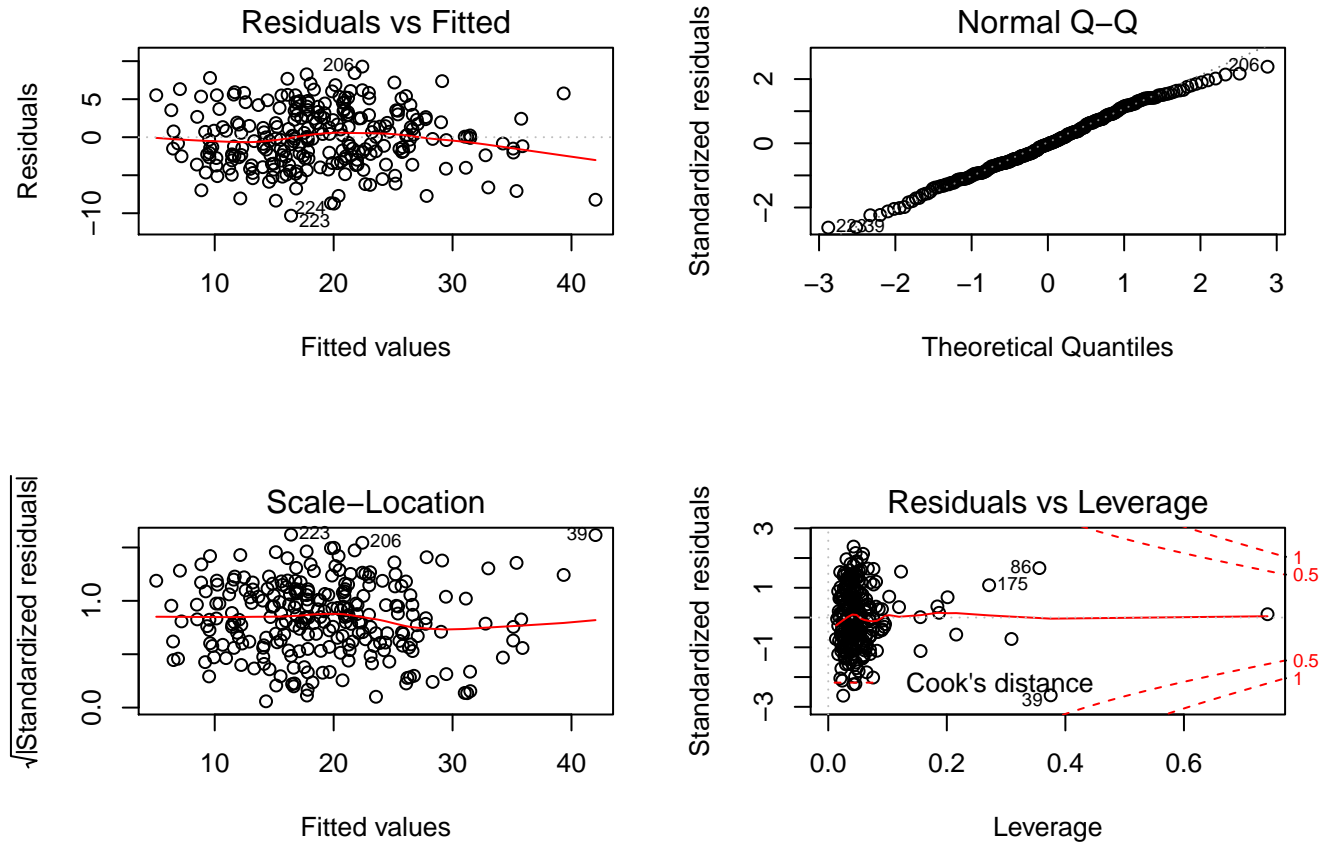
Model diagnostics are shown in Figures 1 and 2

Figure 1: Regression diagnostics for Model (1)

	Model 1
(Intercept)	-12.39 (16.18)
age	0.06 (0.03)
weight	-0.07 (0.05)
height	-0.07 (0.09)
neck	-0.43 (0.21)*
chest	-0.04 (0.09)
abdomen	0.89 (0.08)***
hip	-0.20 (0.13)
thigh	0.21 (0.13)
knee	-0.02 (0.22)
ankle	0.15 (0.20)
bicep	0.17 (0.16)
forearm	0.42 (0.18)*
wrist	-1.49 (0.49)**
R^2	0.74
Adj. R^2	0.73
Num. obs.	251
RMSE	3.98

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Multiple Linear Regression of the Body Fat Data
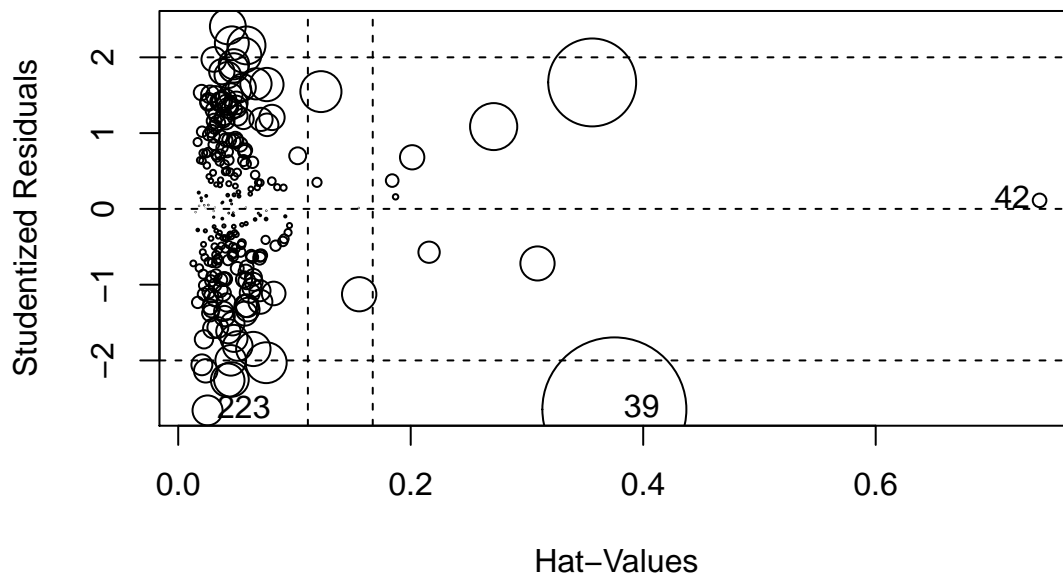
Figure 2: Regression influence plot for Model (1)

Look more closely at observation 42:

pbf1	weight	height
31.70	205.00	29.50

3   Sensitivity Analysis

We perform the same analysis as above, but with observation 42 removed
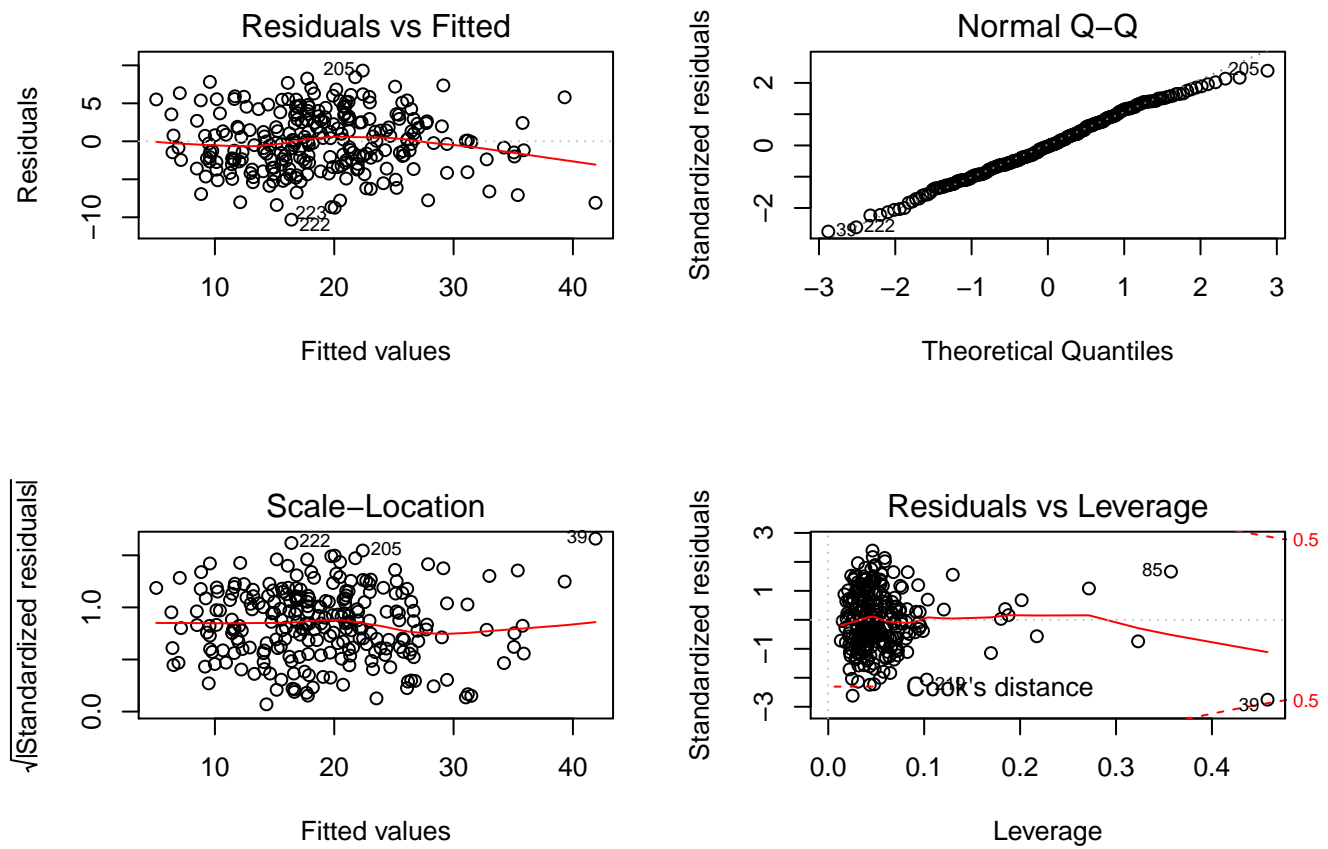
Figure 3: Regression diagnostics for Model (1), with outliers removed

	With obs. 42	Without obs. 42
(Intercept)	$-12.39 \, (16.18)$	$-13.85 \, (20.77)$
age	$0.06 \, (0.03)$	$0.06 \, (0.03)$
weight	$-0.07 \, (0.05)$	$-0.08 \, (0.06)$
height	$-0.07 \, (0.09)$	$-0.06 \, (0.17)$
neck	$-0.43 \, (0.21)^{*}$	$-0.43 \, (0.22)$
chest	$-0.04 \, (0.09)$	$-0.04 \, (0.10)$
abdomen	$0.89 \, (0.08)^{***}$	$0.89 \, (0.08)^{***}$
hip	$-0.20 \, (0.13)$	$-0.20 \, (0.14)$
thigh	$0.21 \, (0.13)$	$0.22 \, (0.14)$
knee	$-0.02 \, (0.22)$	$-0.02 \, (0.23)$
ankle	$0.15 \, (0.20)$	$0.15 \, (0.21)$
bicep	$0.17 \, (0.16)$	$0.17 \, (0.16)$
forearm	$0.42 \, (0.18)^{*}$	$0.42 \, (0.18)^{*}$
wrist	$-1.49 \, (0.49)^{**}$	$-1.49 \, (0.50)^{**}$
R^2	0.74	0.74
Adj. R^2	0.73	0.73
Num. obs.	251	250
RMSE	3.98	3.99

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

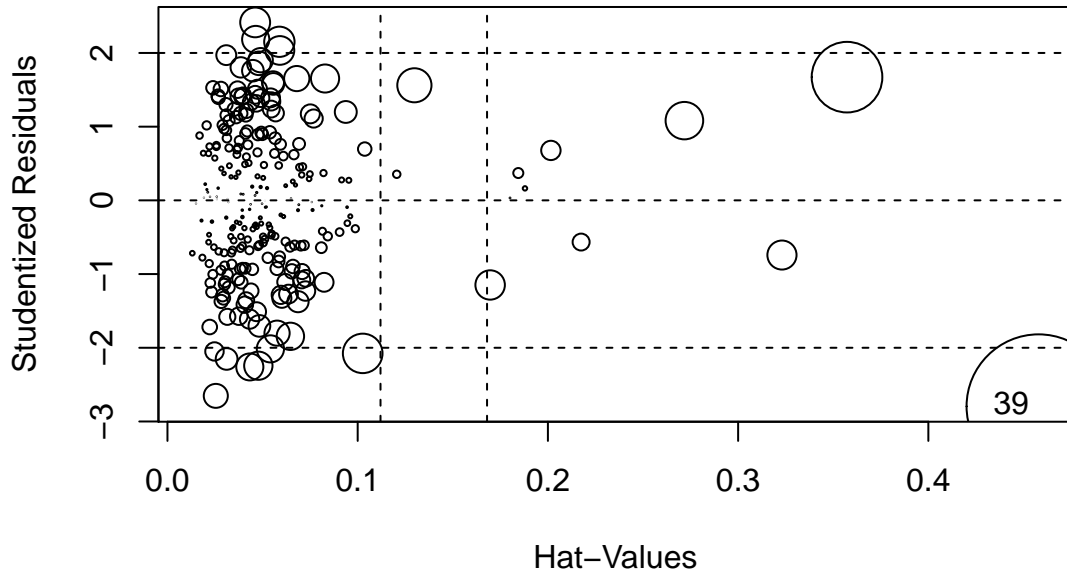Table 2: Sensitivity analsysis; Multiple Linear Regression of the Body Fat Data

Figure 4: Regression influence plot for Model (1), with outliers removed