001-Motivating Example

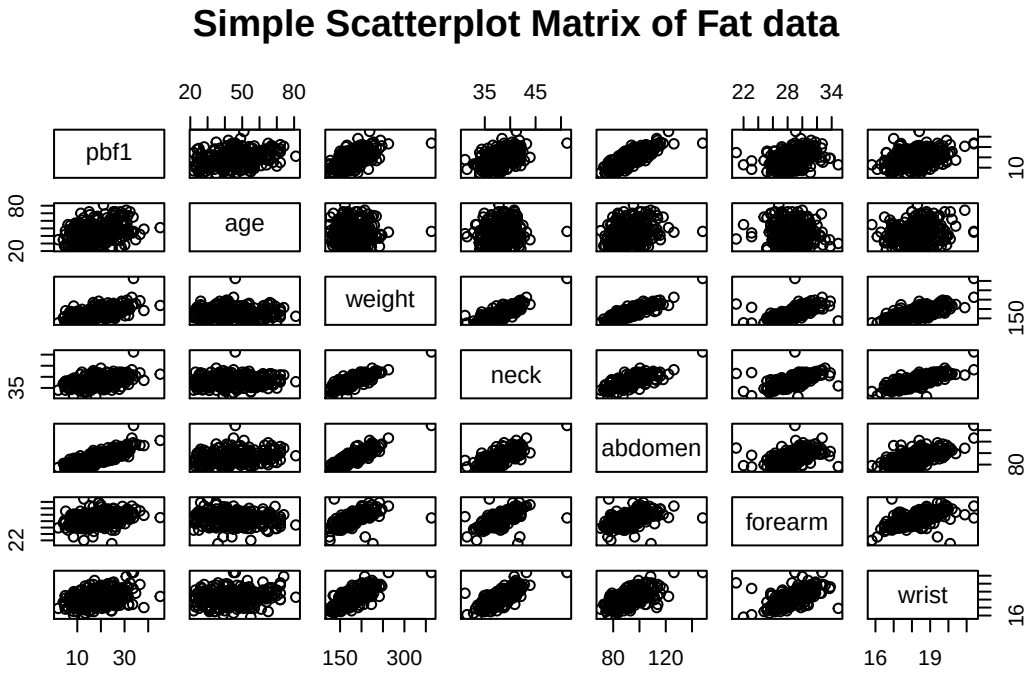
Body Fat Data

May 27, 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 (Penrose et al., 1985; Johnson, 1996; original by Gareth Ambler and modified by Axel Benner, 2015). We model the data using multiple linear regression and perform various model selection techniques.

1 EDA



We will fit a model of the form

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

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

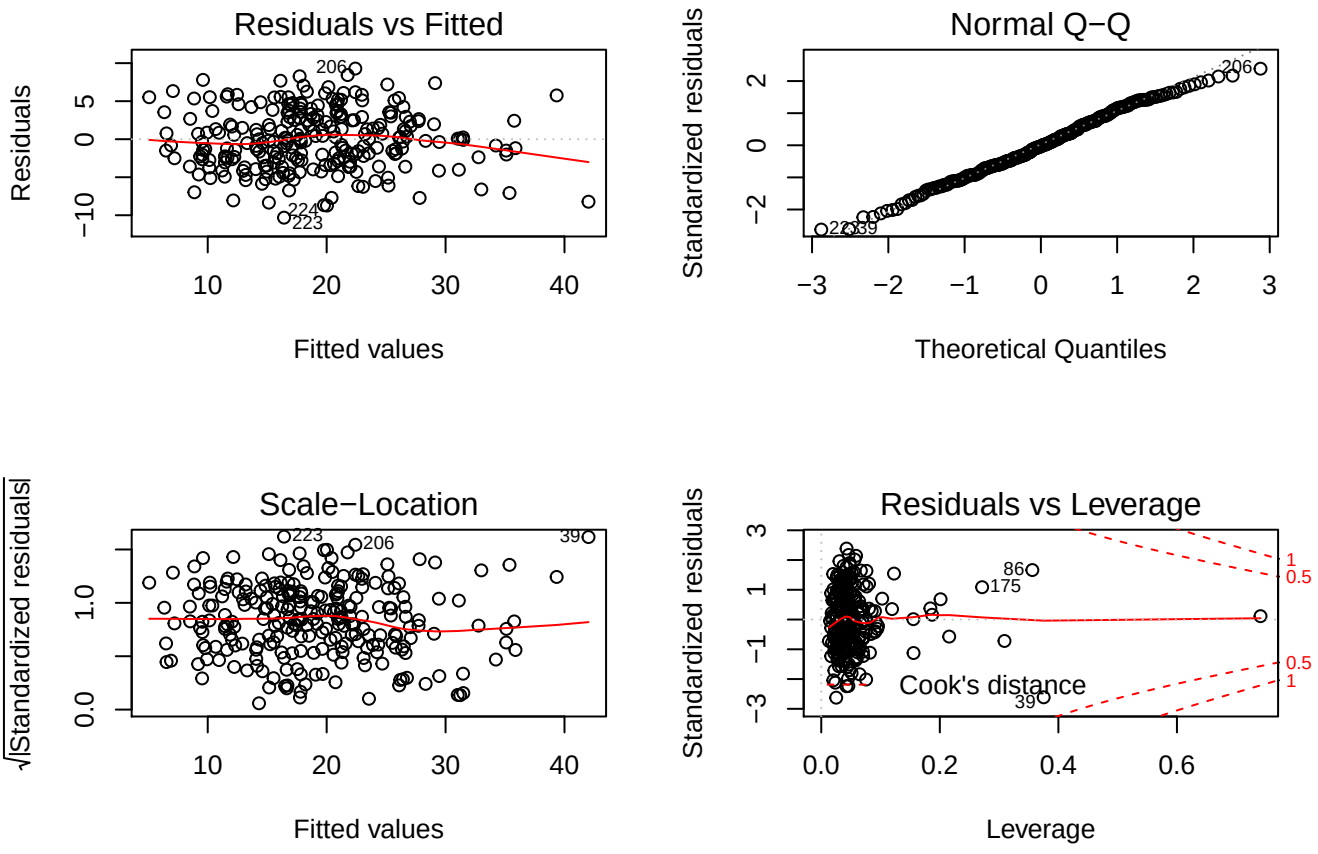


Figure 1: Regression diagnostics for Model (1)

	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

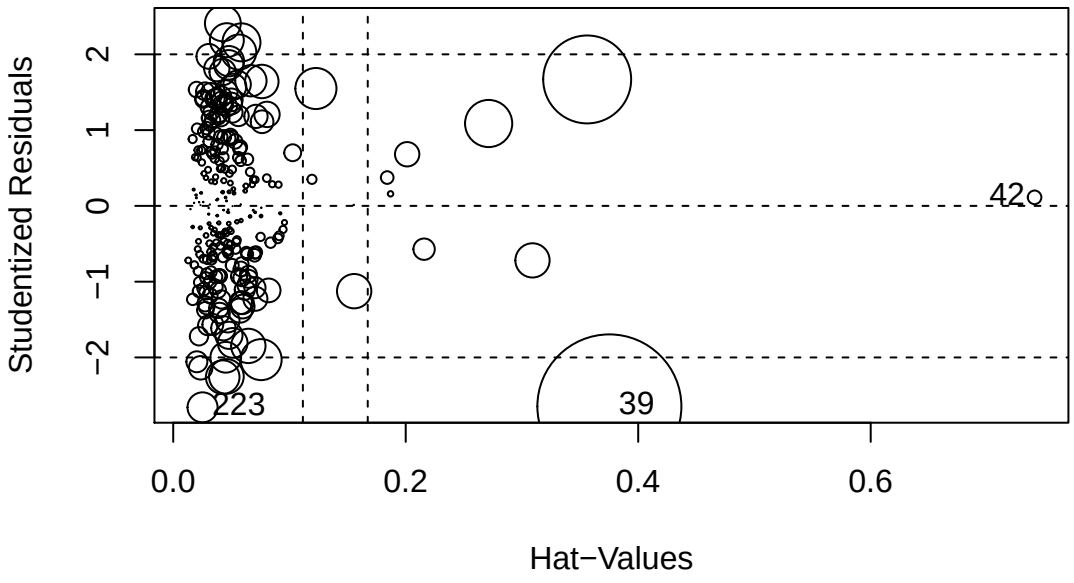


Figure 2: Regression influence plot for Model (1)

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

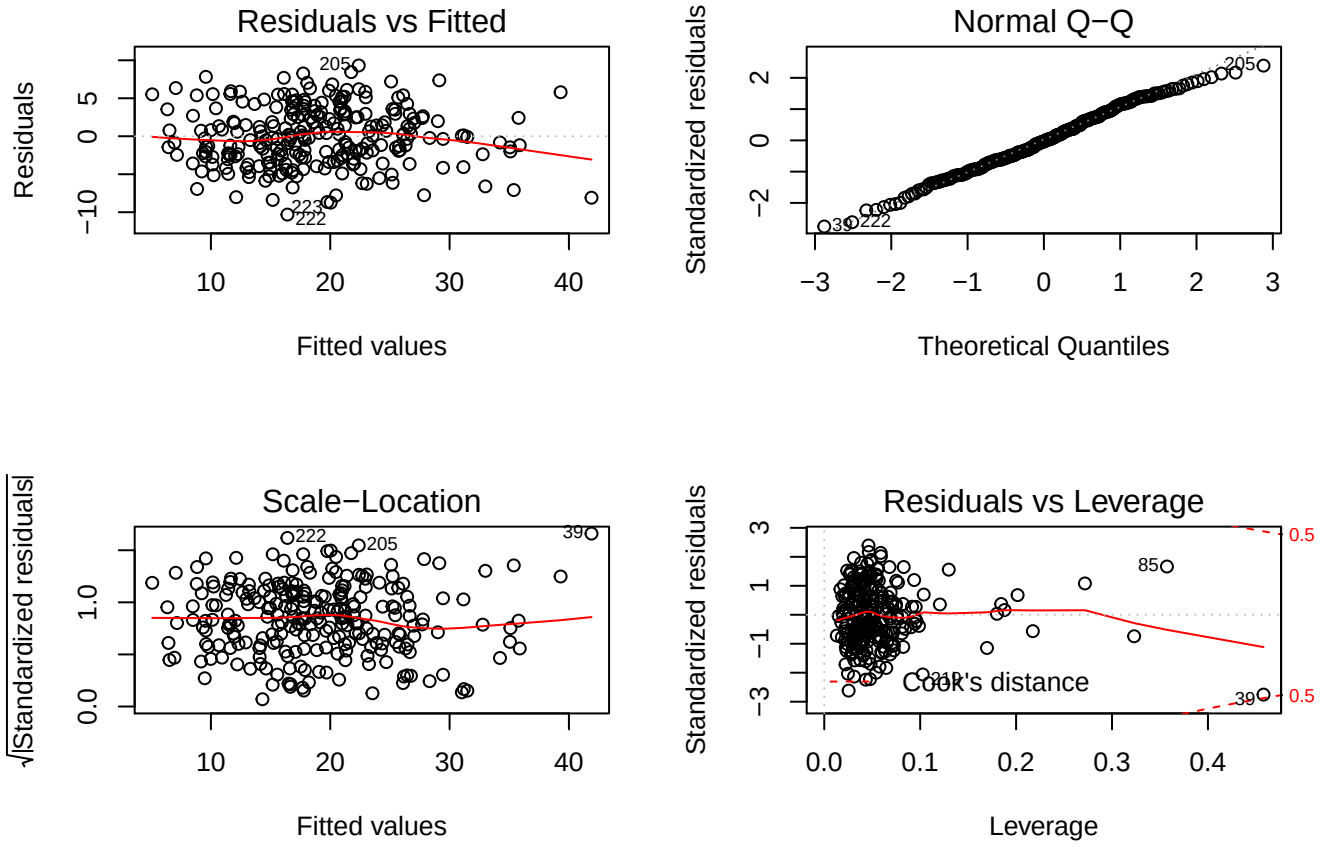


Figure 3: Regression diagnostics for Model (1), with outliers removed

	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

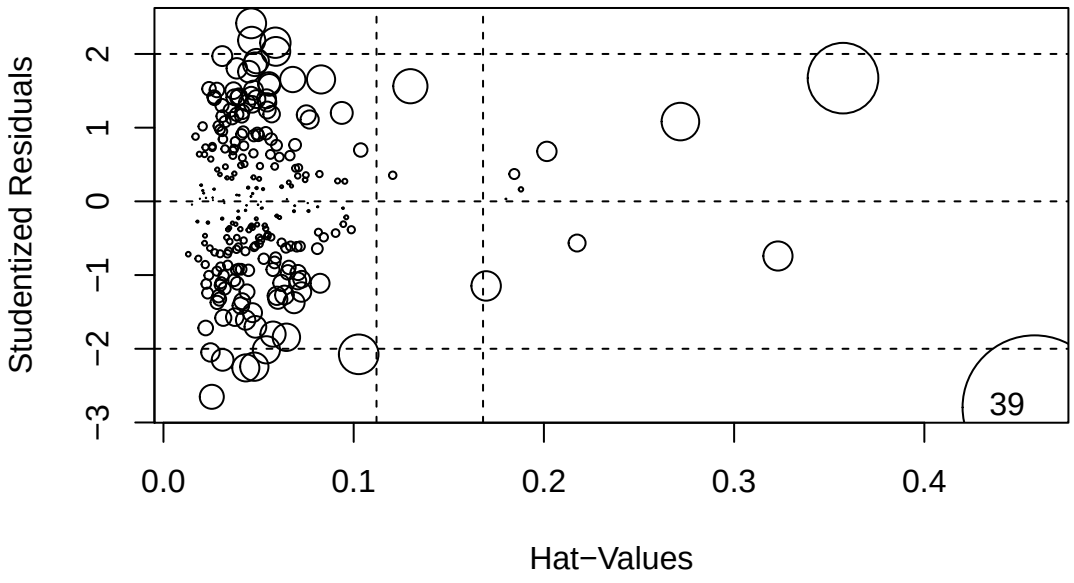


Figure 4: Regression influence plot for Model (1), with outliers removed

References

Roger W Johnson. Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education*, 4(1):265–266, 1996. 1

original by Gareth Ambler and modified by Axel Benner. *mfp: Multivariable Fractional Polynomials*, 2015. URL http://CRAN.R-project.org/package=mfp. R package version 1.5.1. 1

Keith W Penrose, AG Nelson, and A Garth Fisher. Generalized body composition prediction equation for men using simple measurement techniques. *Medicine & Science in Sports & Exercise*, 17(2):189, 1985. 1

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2013. URL http://yihui.name/knitr/. ISBN 978-1482203530.

Yihui Xie. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. URL http://www.crcpress.com/product/isbn/9781466561595. ISBN 978-1466561595.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2015. URL http://yihui.name/knitr/. R package version 1.10.5.

A R Code

```
DT[42, .(pbf1, weight, height), ] %>% xtable %>% print(include.rownames = FALSE)
sessionInfo()
getPckg <- function(pckg) install.packages(pckg, repos = "http://cran.r-project.org")

pckg = try(require(knitr))
if (!pckg) {
    cat("Installing 'knitr' from CRAN\n")
    getPckg("knitr")
    require(knitr)
}

pckg = try(require(data.table))
if (!pckg) {
    cat("Installing 'data.table' from CRAN\n")
    getPckg("data.table")
    require(data.table)
}

pckg = try(require(dplyr))
if (!pckg) {
    cat("Installing 'dplyr' from CRAN\n")
    getPckg("dplyr")
    require(dplyr)
}

pckg = try(require(texreg))
if (!pckg) {
    cat("Installing 'texreg' from CRAN\n")
    getPckg("texreg")
    require(texreg)
}

pckg = try(require(car))
if (!pckg) {
    cat("Installing 'car' from CRAN\n")
    getPckg("car")
    require(car)
}
```

```
pckg = try(require(MASS))
if (!pckg) {
    cat("Installing 'MASS' from CRAN\n")
    getPckg("MASS")
    require(MASS)
}

pckg = try(require(xtable))
if (!pckg) {
    cat("Installing 'xtable' from CRAN\n")
    getPckg("xtable")
    require(xtable)
}
# 1. Percent body fat using Method 1: 457/Density - 414.2 2.
# Age (yrs) 3. Weight (lbs) 4. Height (inches) 5. Neck
# circumference (cm) 6. Chest circumference (cm) 7. Abdomen
# circumference (cm) at the umbilicus and level with the
# iliac crest 8. Hip circumference (cm) 9. Thigh
# circumference (cm) 10. Knee circumference (cm) 11 Ankle
# circumference (cm) 12. Extended biceps circumference (cm)
# 13. Forearm circumference (cm) 14. Wrist circumference (cm)
# distal to the styloid processes
DT <- data.table::fread("fat-data.csv")
pairs(~pbf1 + age + weight + neck + abdomen + forearm + wrist,
    data = DT, main = "Simple Scatterplot Matrix of Fat data")
fit1 <- lm(pbf1 ~ ., data = DT)
texreg::texreg(fit1, digits = 2, caption = "Multiple Linear Regression of the Body Fat Data",
    label = "tab:results", booktabs = TRUE, dcolumn = TRUE, single.row = TRUE,
    use.packages = FALSE)
par(mfrow = c(2, 2))
plot(fit1)
car::influencePlot(fit1)


# Sensitivity Analysis
# -------------------------------------------------
DT <- DT[-c(42), , ]
fit2 <- lm(pbf1 ~ ., data = DT)
```

```
texreg::texreg(list(fit1, fit2), digits = 2, custom.model.names = c("With obs. 42",
    "Without obs. 42"), caption = "Sensitivity analsysis; Multiple Linear Regression of the Body F
    label = "tab:results2", booktabs = TRUE, dcolumn = TRUE,
    single.row = TRUE, use.packages = FALSE)
par(mfrow = c(2, 2))
plot(fit2)
car::influencePlot(fit2)
```

B Session Information

```
sessionInfo()
```

```
## R version 3.2.0 (2015-04-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04 LTS
##
## locale:
##  [1] LC_CTYPE=en_CA.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_CA.UTF-8        LC_COLLATE=en_CA.UTF-8
##  [5] LC_MONETARY=en_CA.UTF-8    LC_MESSAGES=en_CA.UTF-8
##  [7] LC_PAPER=en_CA.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_CA.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets
## [6] methods   base
##
## other attached packages:
## [1] xtable_1.7-4     MASS_7.3-39      car_2.0-25
## [4] texreg_1.35      dplyr_0.4.1      data.table_1.9.4
## [7] knitr_1.10
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.11.6     magrittr_1.5    splines_3.2.0
##  [4] lattice_0.20-31 minqa_1.2.4     highr_0.5
##  [7] stringr_1.0.0   plyr_1.8.2      tools_3.2.0
## [10] nnet_7.3-9      parallel_3.2.0  pbkrtest_0.4-2
## [13] grid_3.2.0      nlme_3.1-120    mgcv_1.8-6
## [16] quantreg_5.11   DBI_0.3.1       lme4_1.1-7
## [19] assertthat_0.1  Matrix_1.2-0    nloptr_1.0.4
## [22] reshape2_1.4.1  formatR_1.2     evaluate_0.7
## [25] stringi_0.4-1   SparseM_1.6     chron_2.3-45
```