001-Motivating Example

Body Fat Data

May 13, 2019

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

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

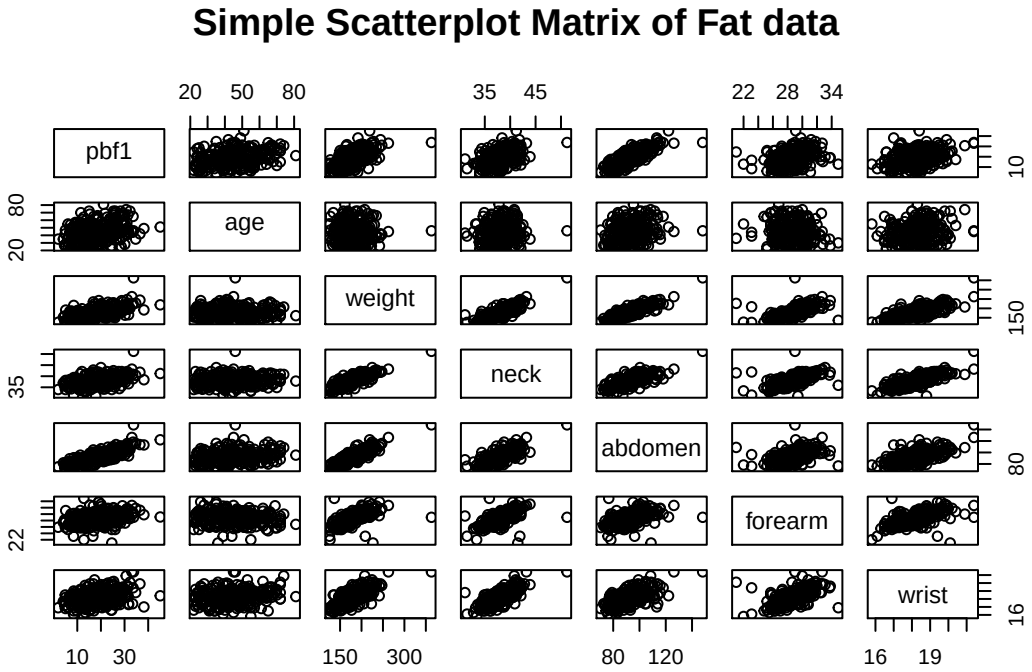


Figure 1: Pairs plot of Body Fat data

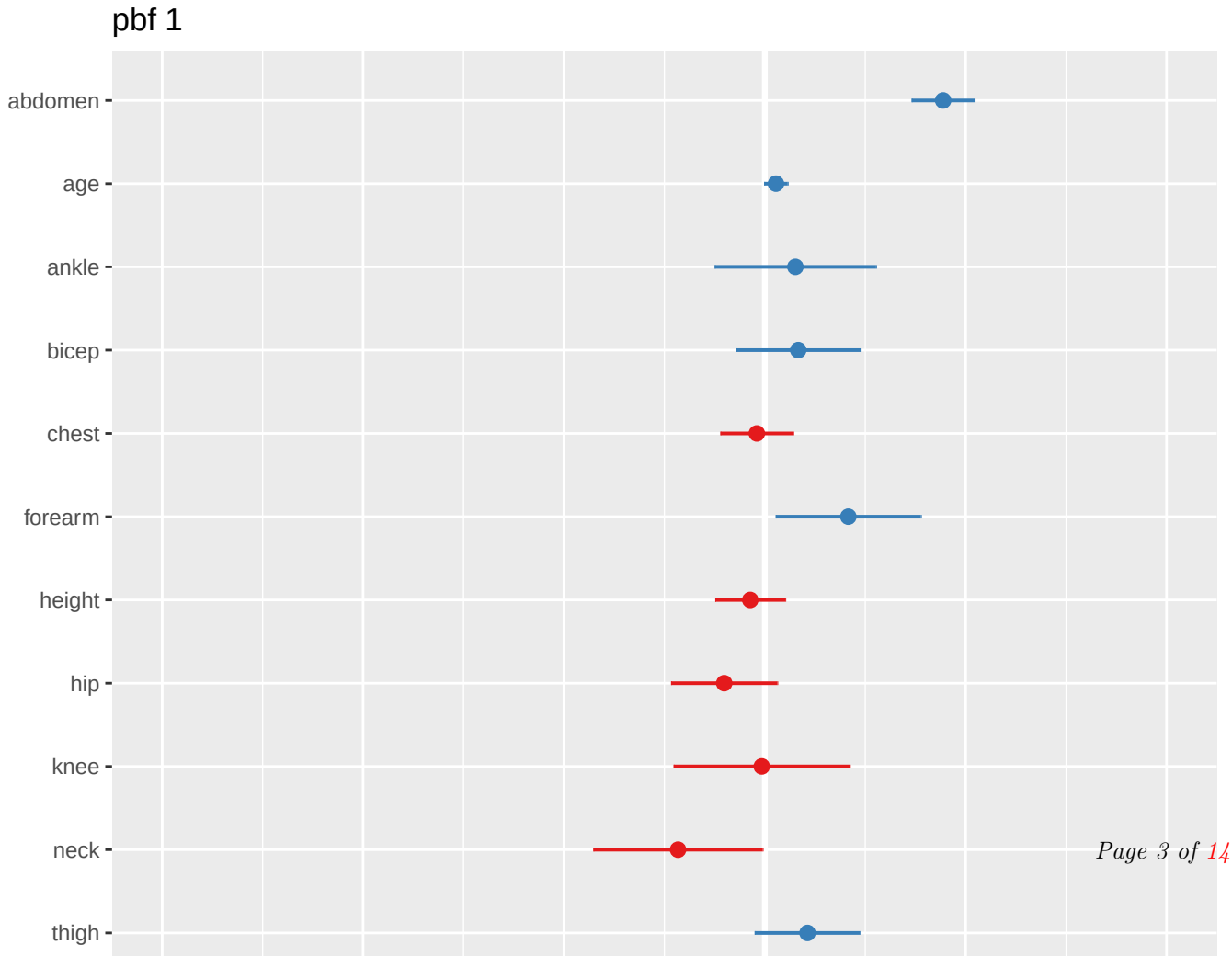
2 Results

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



Model diagnostics are shown in Figures 2 and 3

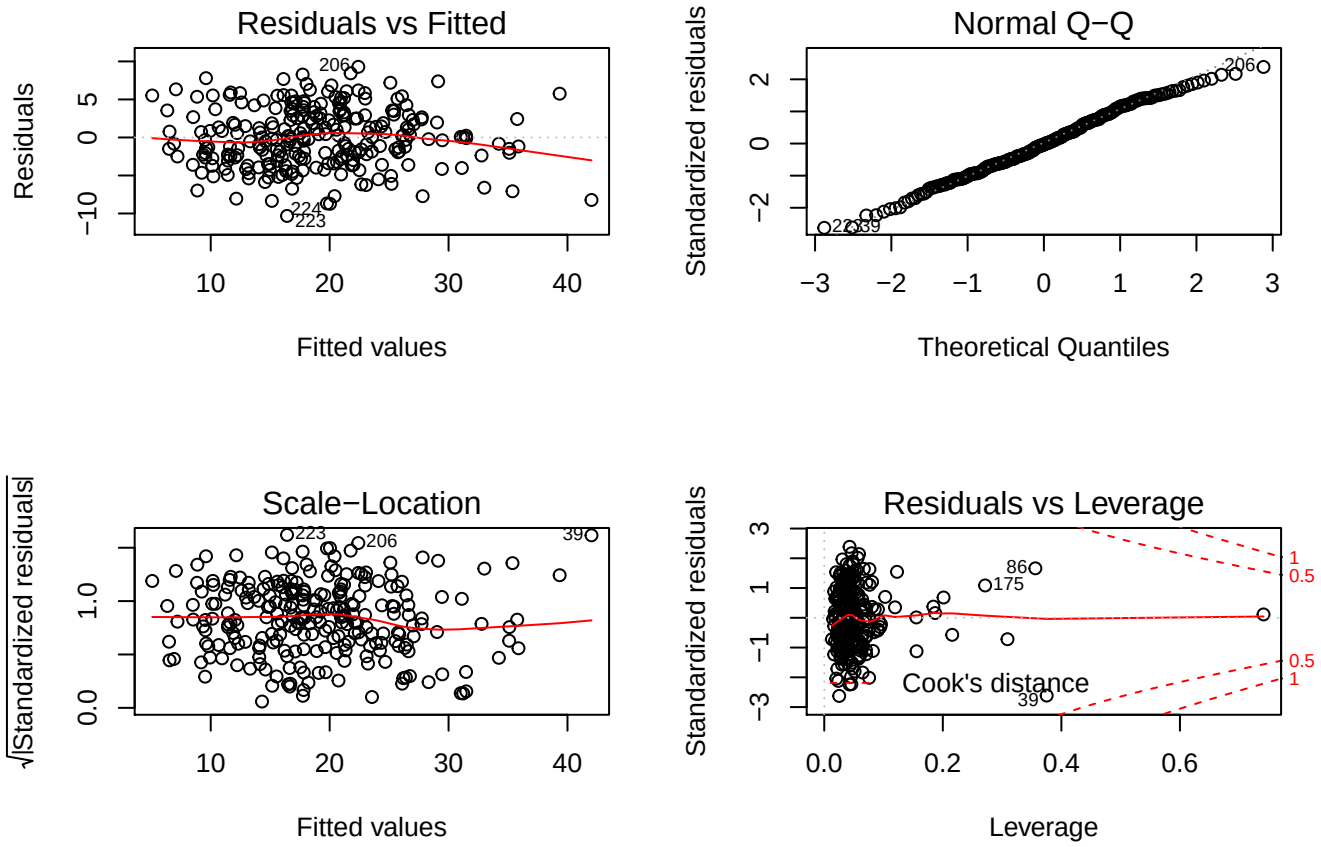


Figure 2: Regression diagnostics for Model (1)

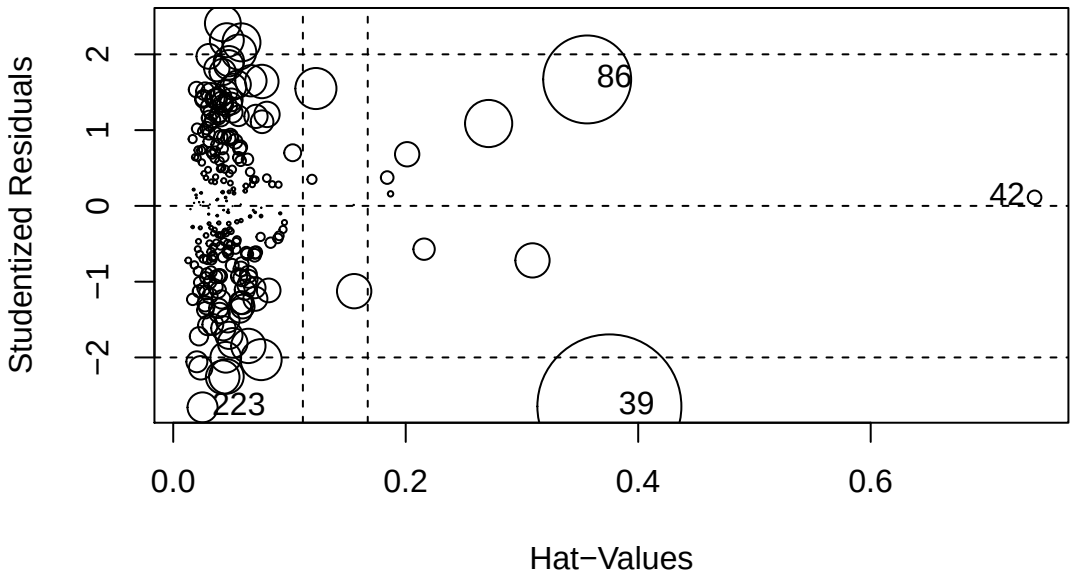


Figure 3: Regression influence plot for Model (1)

Look more closely at observation 42:

pbf1	weight	height
31.7	205	29.5

3 Sensitivity Analysis

We perform the same analysis as above, but with observation 42 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

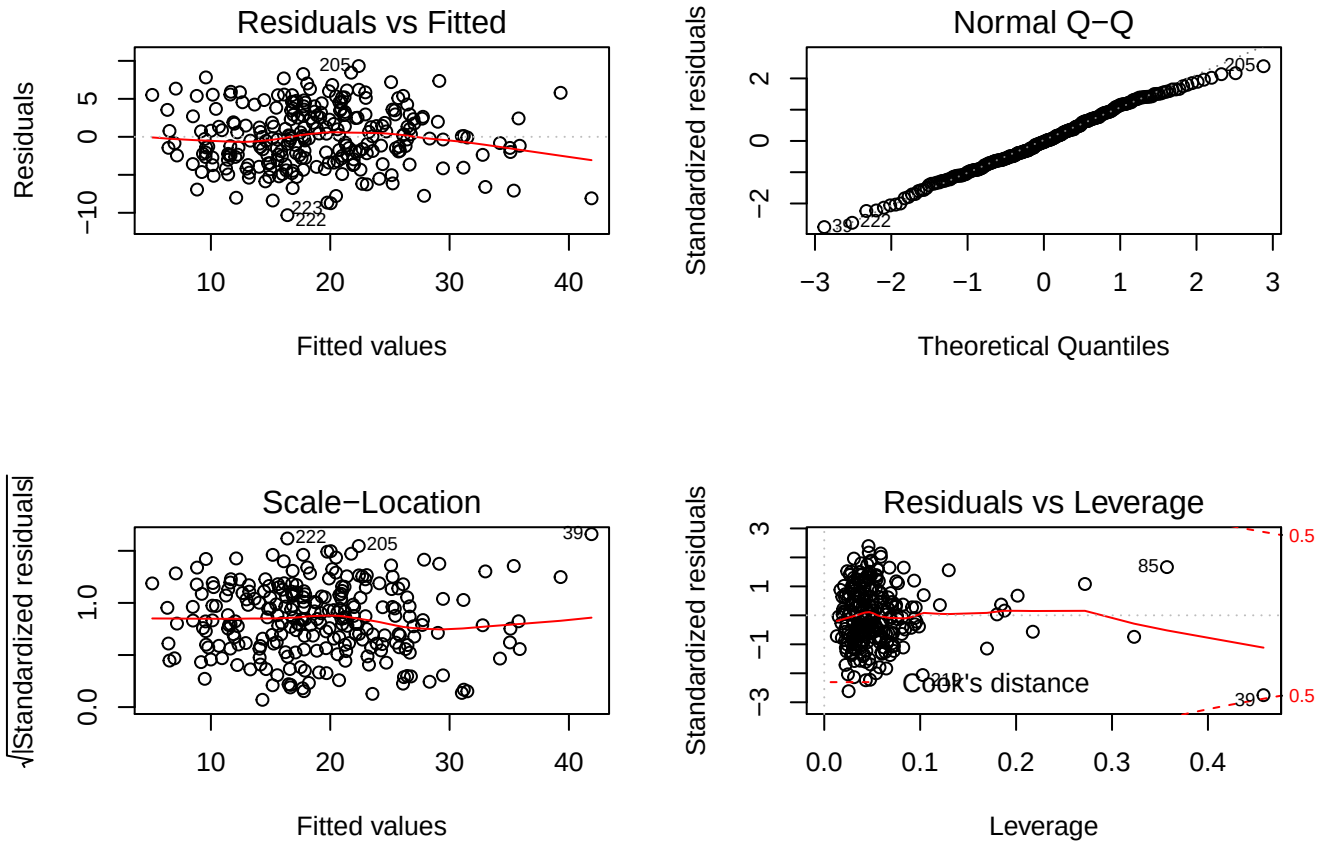


Figure 4: Regression diagnostics for Model (1), with outliers removed

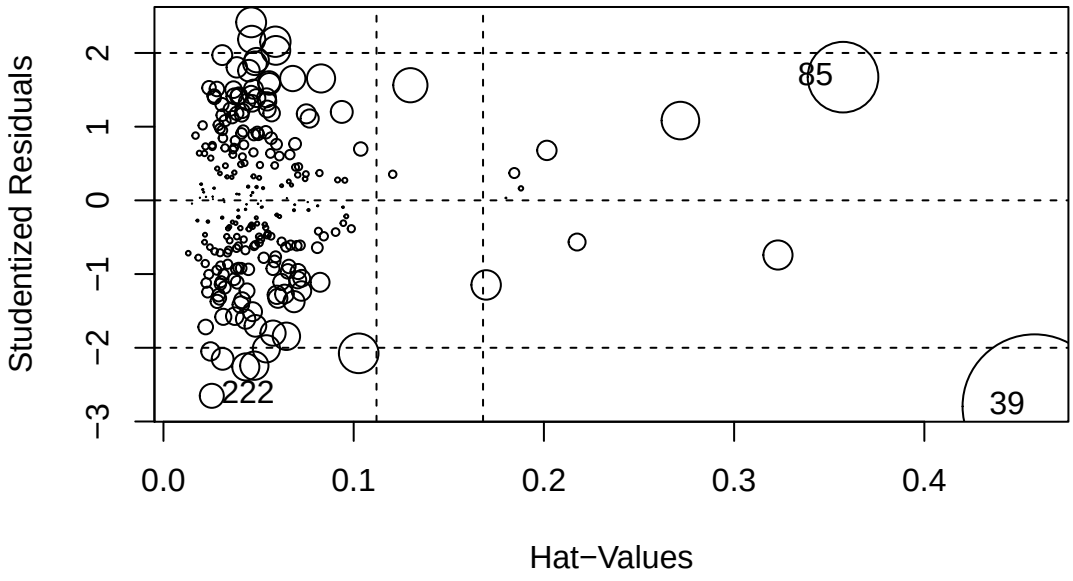


Figure 5: Regression influence plot for Model (1), with outliers removed

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

```
knitr::kable(DT[42,.(pbf1,weight,height),])
print(sessionInfo(), locale=FALSE)
############################################################
# R Source code file for required packages
# for 001-motivating-example.Rmd.
# Hosted on github repo 'sahirbhatnagar/CSSC2018'
# Author: Sahir Bhatnagar
# Created: May 31, 2018
# Updated:
# Notes: we use pacman to install packages. p_load will
# check if the package is installed, install and load it
############################################################
if (!require("pacman")) install.packages("pacman")

pacman::p_load(knitr)

# for easy location of files
pacman::p_load(here)

# data loading and manipulation
pacman::p_load(data.table)

# easy plotting and tables of regression output
pacman::p_load(sjPlot)
pacman::p_load(ggeffects)

# influencePlot
pacman::p_load(car)

# the pipe %>%
pacman::p_load(magrittr)

# this allows you to use bookdown::html_document2 in the YAML
# which allows for figure labelling and referencing
pacman::p_load(bookdown)
############################################################
# R Source code file for loading data
# for 001-motivating-example.Rmd.
```

```
# Hosted on github repo 'sahirbhatnagar/CSSC2018'
# Author: Sahir Bhatnagar
# Created: May 31, 2018
# Updated:
# Notes: This script contains everything related to data
# cleaning. In this example, there isn't any cleaning to do
# but the is NEVER the case in real life consulting.
############################################################



# Description of variables
# 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"
##############################################################################
file_path <- here::here("data", "fat-data.csv")
DT <- data.table::fread(file_path)
############################################################
# R Source code file for analysis
# for 001-motivating-example.Rmd.
# Hosted on github repo 'sahirbhatnagar/CSSC2018'
# Author: Sahir Bhatnagar
# Created: May 31, 2018
# Updated:
# Notes:
############################################################
graphics::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 =
               booktabs = TRUE, dcolumn = TRUE, single.row = TRUE, use.packages = FALSE)
# print results to a table
sjPlot::tab_model(fit1, dv.labels = "Percentage Body Fat")
# plot coefficients
sjPlot::plot_model(fit1, type = "est")
par(mfrow=c(2,2))
plot(fit1)
sjPlot::plot_model(fit1, type = "pred")
car::influencePlot(fit1)
DT2 <- DT[-c(42), , ]
fit2 <- lm(pbf1~., data = DT2)
sjPlot::tab_model(fit1, fit2,
                  dv.labels = c("With obs. 42","Without obs. 42"),
                  p.style = "numeric")
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 Fat Data',
               label = "tab:results2", booktabs = TRUE, dcolumn = TRUE, single.row = TRUE, use.pac
par(mfrow=c(2,2))
plot(fit2)
car::influencePlot(fit2)
```

B Session Information

```
print(sessionInfo(), locale=FALSE)
```

```
## R version 3.6.0 (2019-04-26)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Pop!_OS 18.10
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.8.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.8.0
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] bookdown_0.9      magrittr_1.5      car_3.0-2
##  [4] carData_3.0-2     ggeffects_0.8.0   sjPlot_2.6.2
##  [7] data.table_1.12.0 here_0.1          pacman_0.5.0
## [10] knitr_1.22
##
## loaded via a namespace (and not attached):
##  [1] tidyr_0.8.2         splines_3.6.0       modelr_0.1.4
##  [4] assertthat_0.2.1    highr_0.8           stats4_3.6.0
##  [7] cellranger_1.1.0    coin_1.2-2          pillar_1.3.1
## [10] backports_1.1.3     lattice_0.20-38     glue_1.3.1
## [13] RColorBrewer_1.1-2 glmmTMB_0.2.3       snakecase_0.9.2
## [16] minqa_1.2.4         colorspace_1.4-0    sandwich_2.5-0
## [19] Matrix_1.2-17       plyr_1.8.4          psych_1.8.12
## [22] pkgconfig_2.0.2     broom_0.5.1         haven_2.1.0
## [25] purrr_0.3.2         xtable_1.8-4        mvtnorm_1.0-8
## [28] scales_1.0.0        stringdist_0.9.5.1 openxlsx_4.1.0
## [31] rio_0.5.16          lme4_1.1-20         emmeans_1.3.2
## [34] tibble_2.1.1        bayesplot_1.6.0     generics_0.0.2
## [37] ggplot2_3.1.0       sjlabelled_1.0.16   TH.data_1.0-10
## [40] TMB_1.7.15          lazyeval_0.2.1      mnormt_1.5-5
## [43] readxl_1.3.0        survival_2.43-3     crayon_1.3.4
## [46] estimability_1.3    evaluate_0.13       nlme_3.1-139
## [49] MASS_7.3-51.1       forcats_0.4.0       foreign_0.8-71
## [52] tools_3.6.0         hms_0.4.2           multcomp_1.4-8
```

```
## [55] stringr_1.4.0      munsell_0.5.0      zip_2.0.0
## [58] prediction_0.3.6.2 compiler_3.6.0     rlang_0.3.4
## [61] grid_3.6.0         nloptr_1.2.1       ggridges_0.5.1
## [64] texreg_1.36.23     gtable_0.2.0       codetools_0.2-16
## [67] abind_1.4-5        curl_3.3           sjstats_0.17.3
## [70] sjmisc_2.7.7       R6_2.4.0           zoo_1.8-4
## [73] dplyr_0.8.0.1      pwr_1.2-2          rprojroot_1.3-2
## [76] modeltools_0.2-22  stringi_1.4.3      parallel_3.6.0
## [79] Rcpp_1.0.1         tidyselect_0.2.5   xfun_0.6
## [82] coda_0.19-2
```