Alternative Methods To Estimate Body Fat

Aidan, Ash, Enes, Rajan

DATA2902 T09ol_ontime_1

2020-11-12

Part 1: Introduction

Abstract

- Fat percentage of the human body may be of clinical importance
  - Influence in morbidity and mortality
  - Alter effectiveness of drugs
  - Affect body's ability to retain heat
- Current methods to estimate body fat percentage are expensive and technical
  - Electrical calipers
  - Bioelectrical impedance
  - Hydrostatic weighing

SUBSCRIBE **healthline**

A skin-fold test is done using a tool called calipers to pinch different areas of your body and measure body fat. There are a few ways to measure, but many people go with a three-site approach developed by researchers Jackson and Pollock in the 1980s. This method takes the least amount of time to complete. It's also cost-effective, as you can find calipers online for less than $7.

How to:

- If you're a man, measure fat at your chest, abdominals, and thigh.
- If you're a woman, measure fat at your triceps, suprailiac (about an inch above the hip bone), and the thigh.
- Calipers may come with instructions on how to convert these numbers to your body fat percentage.
- You can also consult an online skinfold calculator if you'd rather not do the math yourself.
- Measure on one side of your body, usually the right, for consistency.
- Mark the pinch site 1 centimeter above the skinfold.
- Consider asking a friend or family member to do the measurements for you.
- Take at least two measurements of the same area and average them for the most accurate data.

Tips

Accuracy

When performed correctly, there is around a +/– 3 percent error rate. You may also do a seven-site measurement. This approach is more time-consuming, but it may be slightly more accurate.

AIM: To estimate body fat percentage within a ~3% margin of error using multiple linear regression on common body measurements

The Data Set

Background

- Collected by the Human Performance Research Center at Brigham Young University
- Originally used to produce predictive equations for body composition (1974)
- Available at https://dasl.datadescription.com/datafile/bodyfat

Variables

Response

- `pct_bf`

General characteristics

- `age` (years)
- `weight` (kg)
- `height` (cm)

Circumference measurements (cm)

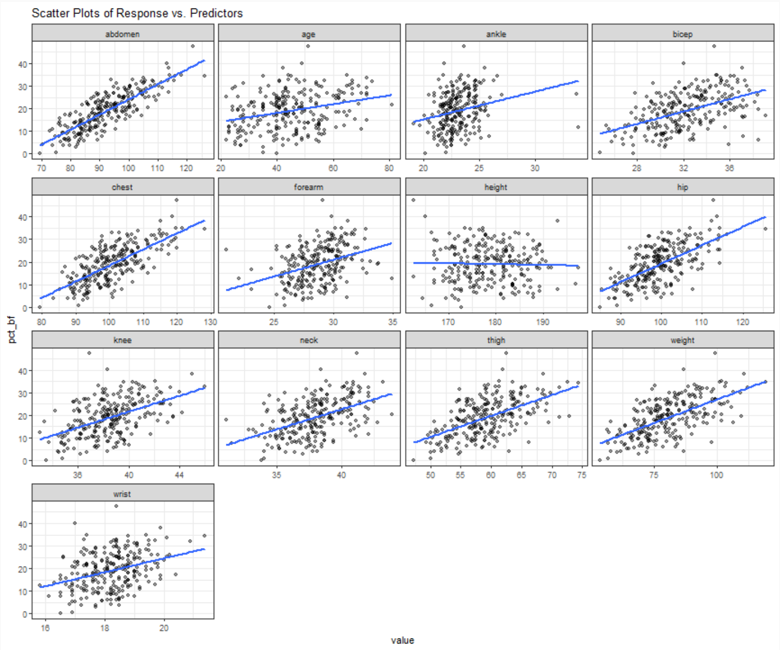
- `neck`, `chest`
- `abdomen`, `hip`
- `thigh`, `knee`
- `ankle`, `bicep`
- `forearm`, `wrist`

Part 2: Analysis

Analysis: Initial Assumptions

☑ Linearity

- Response of `pct_bf` against all predictor variables is linear
- ∴ Linearity assumption is satisfied after EDA



☑ Independence

- All observations independent measurements taken from 250 different males
- ∴ Independence assumption is satisfied

Analysis: Model Exploration

Step-wise search

- Search using AIC and then drop insignificant ($p < 0.05$) predictors

Backward Search

	pct bf	
Predictors	*Estimates*	*p*
(Intercept)	2.90	0.720
age	0.06	**0.020**
height	-0.13	**0.008**
abdomen	0.77	**<0.001**
wrist	-1.91	**<0.001**
Observations	250	
R^2 / R^2 adjusted	0.738 / 0.734	

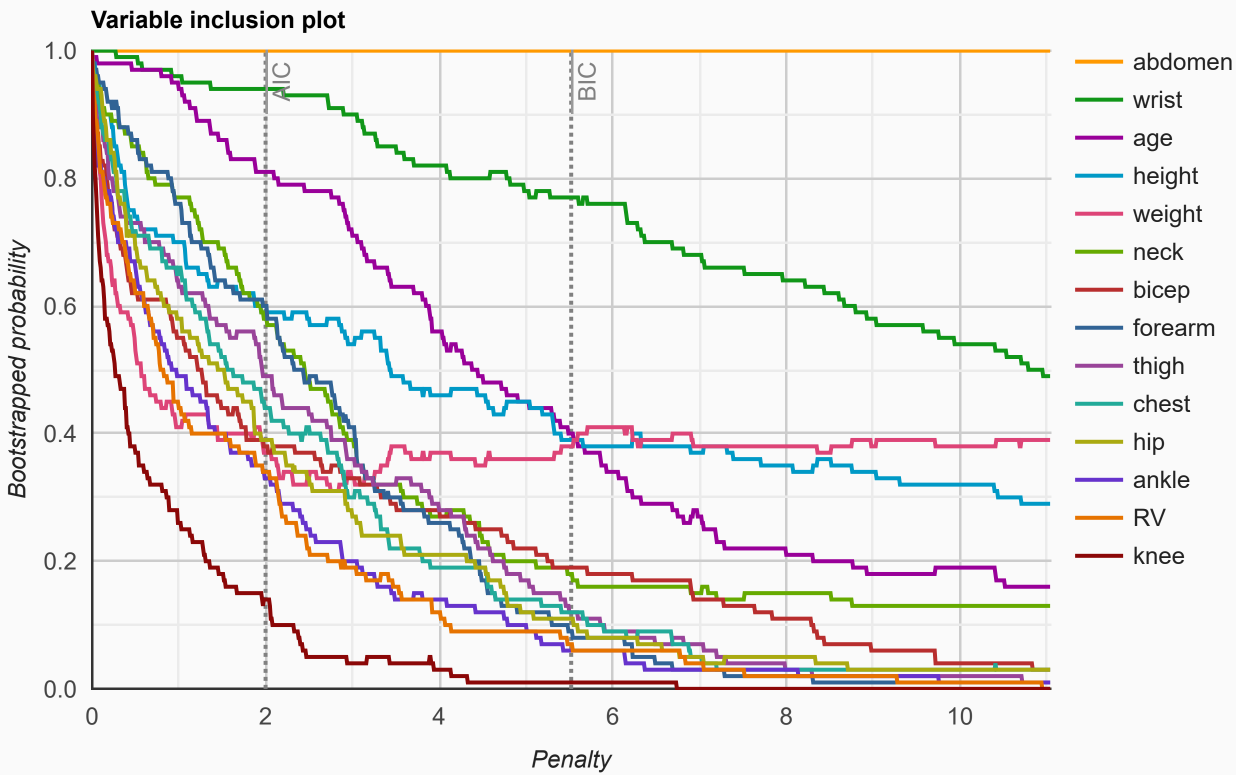
Forward Search

	pct bf	
Predictors	*Estimates*	*p*
(Intercept)	-27.89	**<0.001**
abdomen	0.96	**<0.001**
weight	-0.21	**<0.001**
wrist	-1.37	**0.002**
Observations	250	
R^2 / R^2 adjusted	0.734 / 0.730	

Analysis: Model Exploration

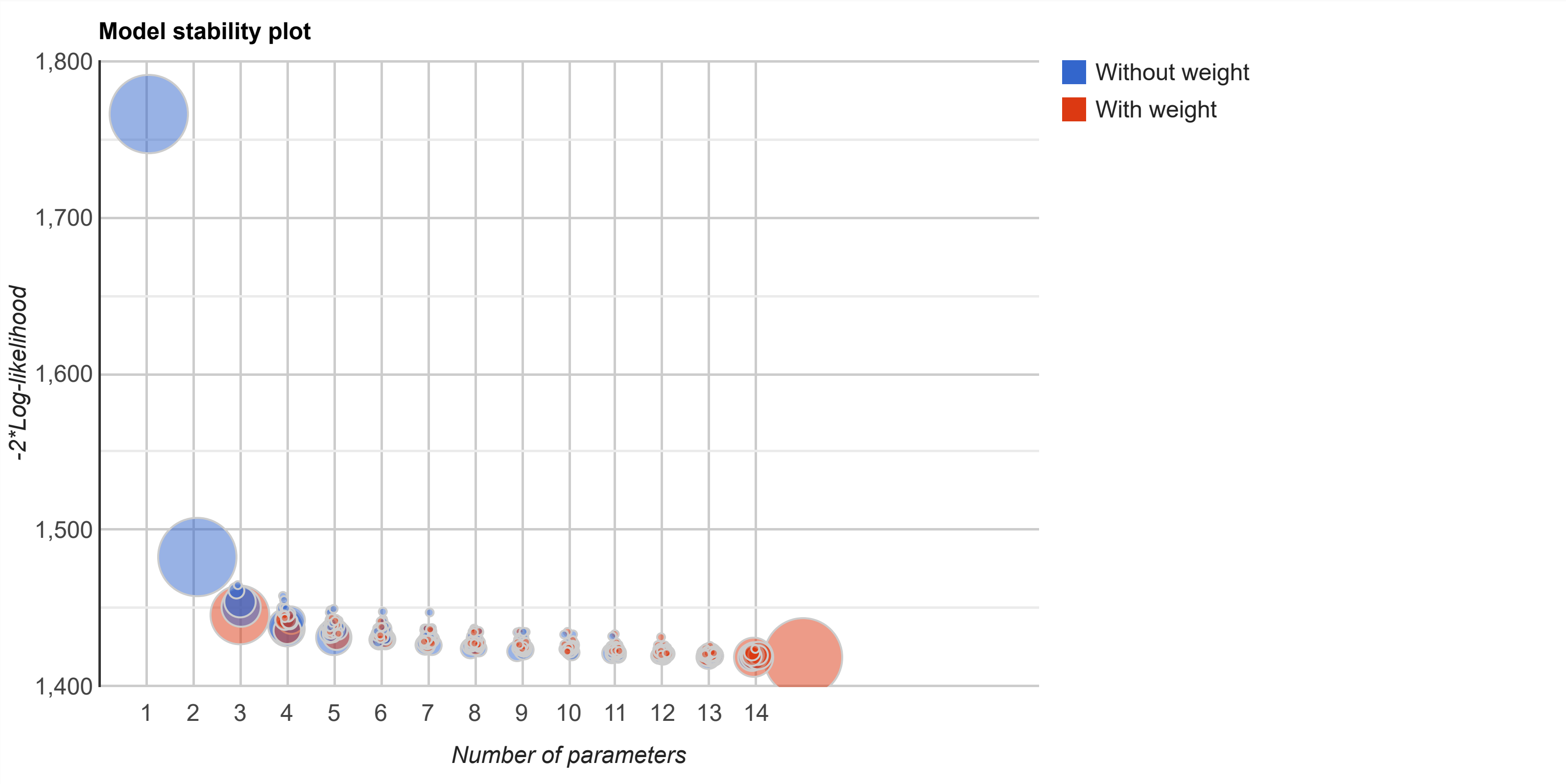
Stability analysis

- Is a certain variable or model consistently selected despite changes in the data set?



Stability analysis

- Is a certain variable or model consistently selected despite changes in the data set?



Analysis: Model Exploration

- Stability plots suggest that these two models are consistent performers:

Abdomen

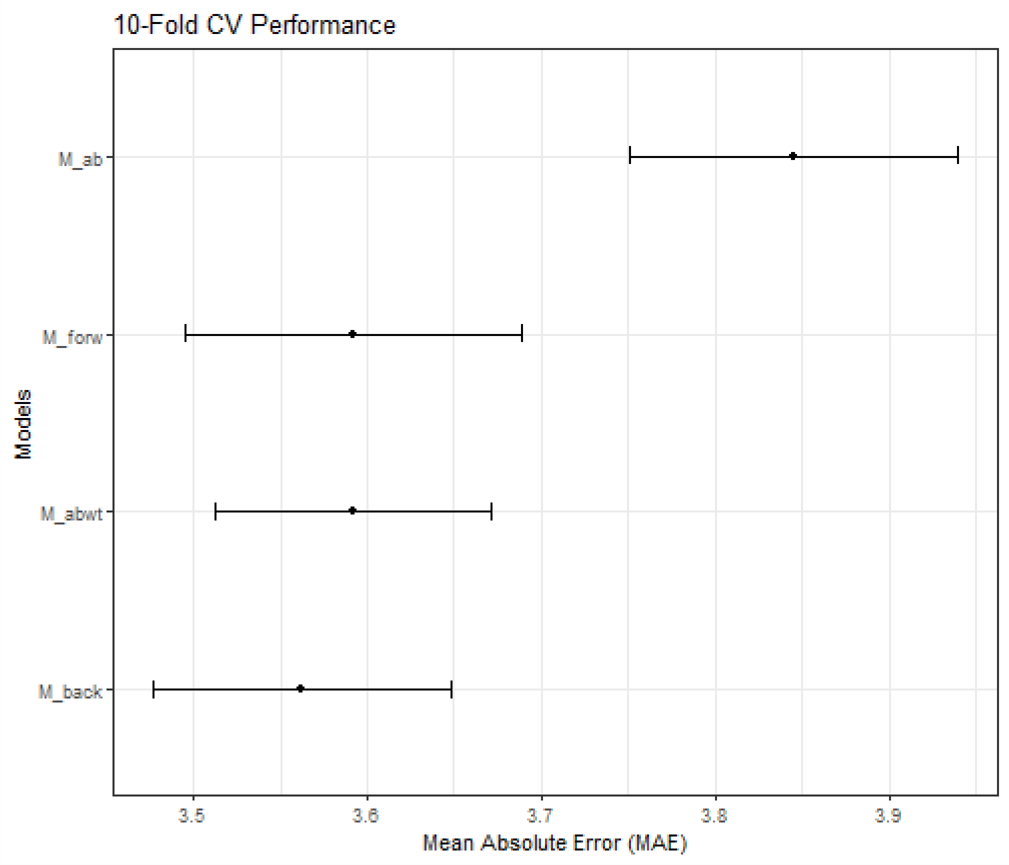
	pct bf	
Predictors	*Estimates*	*p*
(Intercept)	-42.73	**<0.001**
abdomen	0.67	**<0.001**
Observations	250	
R^2 / R^2 adjusted	0.678 / 0.677	

Abdomen + Weight

	pct bf	
Predictors	*Estimates*	*p*
(Intercept)	-47.45	**<0.001**
abdomen	0.98	**<0.001**
weight	-0.29	**<0.001**
Observations	250	
R^2 / R^2 adjusted	0.723 / 0.721	

Analysis: Cross-validation

- Our exploration has yielded 4 models that we would like to compare
- We use 10-fold cross-validation and compare MAEs



Analysis: Model Selection

- We choose `M_abwt`, which is specified as:

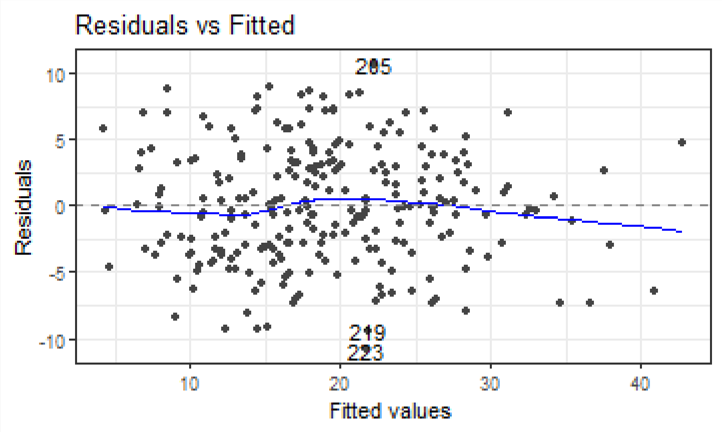
$$\widehat{\text{pct\_bf}} = -47.45 + 0.98 \text{ abdomen} - 0.29 \text{ weight}$$

- Good CV performance
- Parsimonious
- Highly significant coefficients
- Very stable

Analysis: Final Assumptions

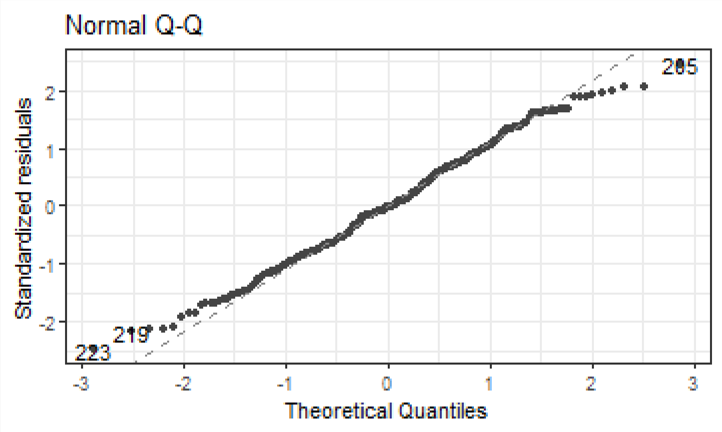
☑ Equal variance

- Residuals vs fitted plot shows no clear patterns
- Fairly equal distribution of points above and below zero line
- Homoskedasticity (equal variance) assumption satisfied



☑ Normality

- Residuals lie mostly along QQ line
- Large amount of observations, $n = 250$
- $\therefore$ CLT will apply
- Hence normality assumption satisfied



Part 3: Results

Results

$$\widehat{\text{pct\_bf}} = -47.45 + 0.98\ \text{abdomen} - 0.29\ \text{weight}$$

Abdomen + Weight

	pct bf	
Predictors	*Estimates*	*p*
(Intercept)	-47.45	**<0.001**
abdomen	0.98	**<0.001**
weight	-0.29	**<0.001**
Observations	250	
R^2 / R^2 adjusted	0.723 / 0.721	

Performance

- CV MAE of ~3.6
- R^2 of 0.723

Interpretations

- +1cm in abdomen circumference results in +0.98% in body fat on average, holding all else constant
- +1kg in weight results in -0.29% in body fat on average, holding all else constant

Inferences

- Abdomen circumference and weight are both highly significant predictors

Part 4: Conclusion

Conclusion

Link to abstract

- Cross-validation score shows that we were on average ~3.6% off the true body fat percentage
- Very similar accuracy to the calipers method
- Easier to measure abdomen circumference and weight

Assumptions and limitations

- Our model met the linearity and $\varepsilon \sim N(0, \sigma^2)$ assumptions
- The source of the data (1974 study) suggested that independence was satisfied

Future research

- Try other regression methods, e.g. kNN and LASSO
- Implement a train/validation/test workflow to select among these methods (may require more data)

References

Data set

- Datadescription.com. (2019). Bodyfat | DASL. [online] Available at: https://dasl.datadescription.com/datafile/bodyfat.

Healthline article

- Healthline. (2018). How to Measure Body Fat: 6 Methods to Try. [online] Available at: https://www.healthline.com/health/how-to-measure-body-fat#calipers.

Packages

- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Yihui Xie (2020). xaringan: Presentation Ninja. R package version 0.18.1. https://github.com/yihui/xaringan
- Tarr G, Müller S, Welsh AH (2018). "mplot: An R Package for Graphical Model Stability and Variable Selection Procedures." *Journal of Statistical Software*, *83*(9), 1-28. doi: 10.18637/jss.v083.i09 (URL: https://doi.org/10.18637/jss.v083.i09).

References

- Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret
- Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." The R Journal 8.2 (2016): 478-489.
- Lüdecke D (2020). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.6, https://CRAN.R-project.org/package=sjPlot

Code help

- Grolemund, Y.X., J.J. Allaire, Garrett (n.d.). Chapter 7 xaringan Presentations | R Markdown: The Definitive Guide. [online] bookdown.org. Available at: https://bookdown.org/yihui/rmarkdown/xaringan.html.
- Grolemund, Y.X., J.J. Allaire, Garrett (n.d.). 3.1 HTML document | R Markdown: The Definitive Guide. [online] bookdown.org. Available at: https://bookdown.org/yihui/rmarkdown/html-document.html.
- Grolemund, G. and Hadley Wickham (2017). R for Data Science. [online] Had.co.nz. Available at: https://r4ds.had.co.nz/index.html.
- Stack Overflow. (2019). Stack Overflow - Where Developers Learn, Share, & Build Careers. [online] Available at: https://stackoverflow.com/.