Body Fat Percentage Model

STAT 628 Module 2 - Group 6
Yifan Zhang, Chenyu Jiang, Amy Merkelz

Motivation

- **Background:** Body fat percentage can be used as a metric for overall health and fitness
- **Problem:** High-accuracy estimates of body fat percentage tend to be complex to measure or require specialized tools:
  - Skin-fold calipers
  - Measuring body density via water displacement
- **Goal:** Determine a model for estimating body fat percentage in men that is:
  - **Simple**
  - **Accurate**
  - **Robust**

Data Cleaning

- **Imputation:**
  - Height : We impute due to mismatches between data and BMI formula

Individual (IDNO)	Original Obs. (m)	Imputed Obs.(m)	Imputation Method
42	**0.74**	**1.76**	**BMI = WEIGHT(kg)/HEIGHT(m)^2**

  - Body Fat : We impute due to the laws of the human body

Individual (IDNO)	Original Obs. (%)	Imputed Obs.(%)	Imputation Method(where SEX is 0 for female and 1 for male)
182	**0**	**14.72**	**Body fat=(1.2 x BMI)+(0.23 x AGE)-(10.8 x SEX)-5.4**

- **Final Cleaned Data:** n = 252 with p = 14 predictors
  - Predictors : ADIPOSITY, HEIGHT, WEIGHT, NECK, CHEST, ABDOMEN, HIP, THIGH, KNEE, ANKLE, BICEPS, FOREARM, WRIST

Model Trade-offs

- **Simplicity:** Linear models with no more than 2 parameters
  - Simplicity scoring will favor 1-parameter models

- **Accuracy:** Correlation Coefficient (R^2)
  - Measures strength of linear relationship between body fat percentage and the parameter(s)

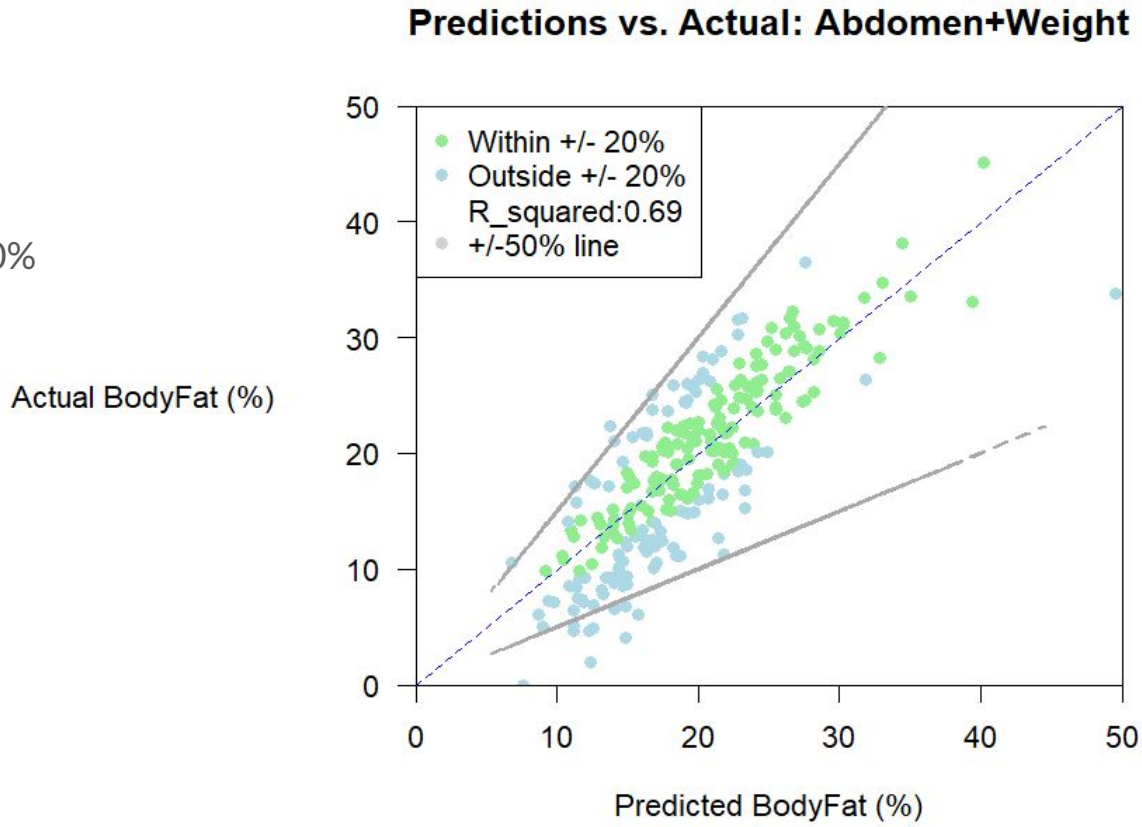
- **Robustness:** Check sum of squares for 4 points against a prior model: US Navy Method
  - US Navy method uses abdomen, height, and neck measurements
  - Calculator used: https://www.calculator.net/body-fat-calculator.html

Model Evaluation

Model	R^2	Number of Parameters	Sum of Squares with US Navy Method
ABDOMEN	**0.67**	**1**	**16.78**
ABDOMEN + WEIGHT	**0.69**	**2**	**7.98**
ABDOMEN + HEIGHT	**0.67**	**2**	**16.63**

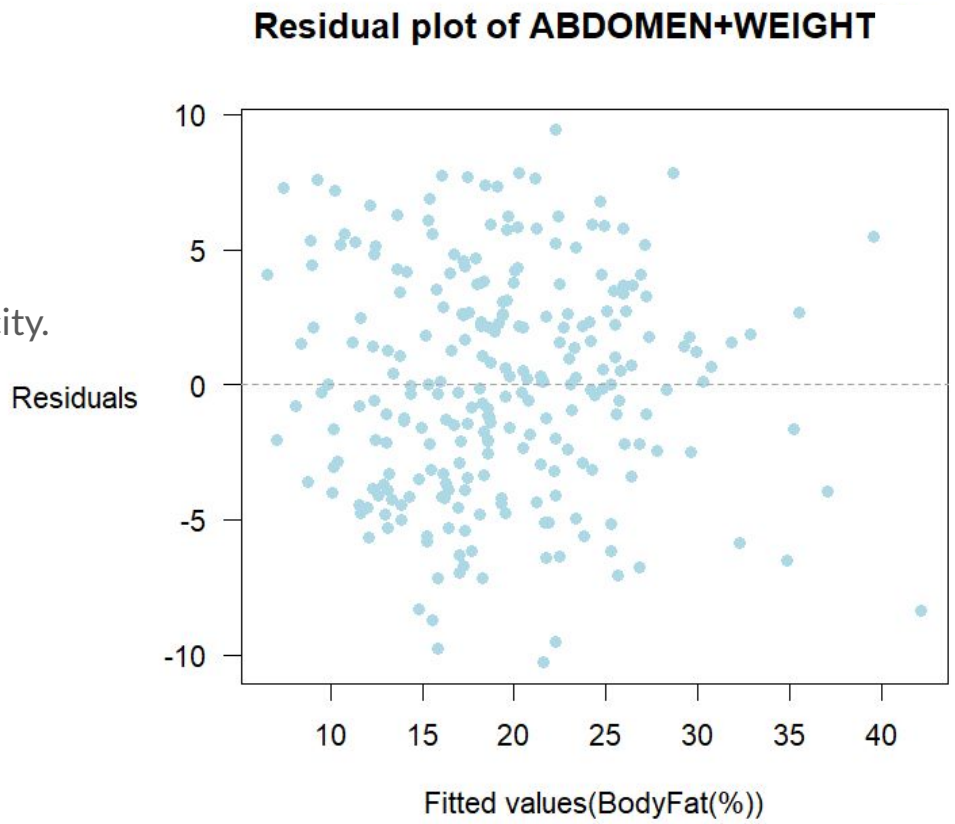
Observations

- Number of points within 20% of the true value: 147/252



Observations

- No clear pattern in residuals:
  supports linearity and homoscedasticity.



Model Strengths and Weaknesses

Final Model: Body Fat ~ -36.92+ 0.88*ABDOMEN-0.31*WEIGHT

Strengths

- Explains 69% of the variation in body fat percentage.
- Simplicity: Only 2 predictors
- Robustness: Lowest Sum of Squared Differences compared with prior model & residual plots

Weaknesses

- Prediction accuracy:
  - Only 58% of predictions fall within ±20% of the true value.
  - Only 91% of predictions fall within ±50% of the true value.

Shiny App Demo

https://amerkelz.shinyapps.io/module2_app/

References

Bailey, Covert (1994). Smart Exercise: Burning Fat, Getting Fit, Houghton-Mifflin Co., Boston, pp. 179-186.

Durnin, J. V. G. A., and Womersley, J., (1974). “Body Fat Assessed from Total Body Density and Its Estimation from Skinfold Thickness: Measurements on 481 Men and Women Aged from 16 to 72 Years.” British Journal of Nutrition 32.1 pp. 77–97. Web.

Siri, W.E. (1956), "Gross composition of the body", in Advances in Biological and Medical Physics, vol. IV, edited by J.H. Lawrence and C.A. Tobias, Academic Press, Inc., New York.

“Body Fat Calculator.” Calculator.Net, www.calculator.net/body-fat-calculator.html. Accessed 16 Oct. 2024.

Akinwande, M. , Dikko, H. and Samson, A., (2015). “Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis.” Open Journal of Statistics, 5, pp. 754-767. Web.

Thank you!

Appendix

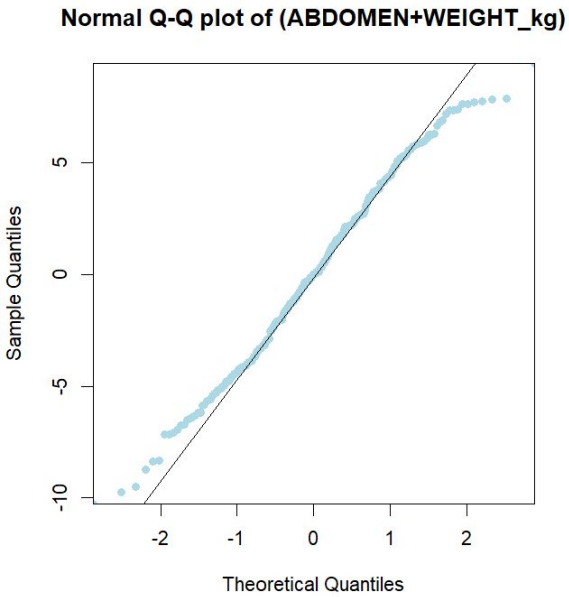
Model	ABDOMEN+WEIGHT	ABDOMEN+HEIGHT
VIF value	4.73	1.04

Table 1: VIF values for the 2 factor models

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}$$

where $R^2{}_i$ is the coefficient of determination of the regression equation in step one, with X_i on the left hand side, and all other predictor variables (all the other X variables) on the right hand side.

Appendix


Normal Q-Q plot of (ABDOMEN+WEIGHT_kg)
Sample Quantiles
5
0
-5
-10
-2
-1
0
1
2
Theoretical Quantiles