Linear regression model for predicting percent body fat

First Author

This version was compiled on November 5, 2022

Per cent body fat, although an accurate health indicator, is difficult and costly to measure. We ran a series of multiple regression analyses with the aim of creating a model to accurately estimate a male's percentage of body fat from a few easily obtained body measurements. Our final model was selected using the backwards stepwise method, with BIC as our threshold indicator. The model contains the predictor variables: age, height, abdomen and wrist circumference.

Introduction. Obesity has become an increasingly important global health problem (Jaacks *et al.*, 2019). It is defined as the accumulation of excess fat in the body, to the extent of impacting an individual's health and well-being (Deurenberg and Yap, 1999). Per cent body fat is a measurement that can help define individuals into a weight category, such as obese (Trang *et al.*, 2019). Furthermore, per cent body fat can be used to as a heath and fitness indicator. However, per cent body fat is costly and difficult to accurately measure (Deurenberg *et al.*, 2002). Hence, the aim of our analysis is to create a model that estimates a male's per cent body fat from easily accessible measurements.

Dataset description. Our data set contains the measurements of 250 men of various ages (Penrose *et al.*, 1985). The per cent body fat of each male was estimated from their body density. In addition, measures such as age, height, weight, and the circumference of the neck, abdomen and chest were taken. In total there are 15 measurements, all classified as continuous, apart from age which is discrete. For data cleaning, we removed the density variable and two observations with less than one per cent body fat. In terms of data wrangling we converted the units for height and weight into the metric system and calculated BMI for the remaining 248 observations (refer to Appendix I).

Model selection. For our model selection, the response variable is body fat percentage and the other measurements including BMI are the potential predictor variables.

Assumption checking.

- Linearity Through looking at the plot of the residuals against the fitted , we can see a fairly horizontal line to 0 however it's a little worrying near the end portion on the line where it seems to develop a sort of declining pattern. This inspired a further look into the assumption through looking at each predictor variable plotted against body fat percentage individually (Figure _) which seems to demonstrate linearity as points above and below the line are fairly equal in proportion.
- Independence Independence can't really be tested through the data, we just placed trust in the fact that the data was collected independently of each other allowing errors to be independent. This assumption should be satisfied at the data comes from the Brigham Young University which is a fairly trusted source, they state the data comes from 250 men at varying ages hopefully indicating independence.
- Normality In the normality assumption we are assuming the residual error is normally distributed. Judging from the qq-plot generated from the residuals (Figure _), it seems to follow the line fairly well up until the end portion, which again was a little worrying. However even if qq-plot doesn't demonstrate normality we should be able to argue the Central Limit Theorem allowing for this assumption due to the large number of observations.
- Homoscedasticity: Through plotting the standardized residuals and fitted values we should be able to test for homoscedasticity where residuals are assumed to have a constant variance. In order to have this assumption the points should be equally spread above and below the line as well as having a line horizontal (or void of any patterns). The graph shows equally spread points and the line isn't horizontal though it seems void of patterns enough to be able to assume homoscedasticity.

Results. **Model equation**

$$bodyfat\% = 0.058*Age - 11.274*Height + 0.760*Abdomen - 1.842*Wrist$$

Our final model is obtained from a backward step wise linear regression model with BIC as selection criteria. This is the fittest among all our models with only 4 predictors.

type	MAE	RMSE
Forward AIC	3.55	4.25
Backward AIC	3.52	4.28
Backward AIC without intercept	3.52	4.25
BIC without intercept	3.52	4.24

Discussion.

- Effectiveness

Our model has a total of four variables, each of which has a p-value less than 0.05. This represents a significant effect of each of the variables on the model.Our adjusted R square value is 0.9586.This represents that our model has a high predictive accuracy. Our model can be used by back health organizations to measure the healthiness of people's body fat percentage.

- Limitation

We used only physical data as a reference. This leads to a large error when dealing with special populations. For example, fit elderly people, people who are short in height but work out regularly and people of different races.. Future study to overcome limitations:We consider increasing the number of elements including daily calorie intake, time spent in daily exercise, etc. to reduce the effect of single body data on body fat percentage.

Conclusion. TBC

Acknowledgments. This template package builds upon, and extends, the work of the excellent rticles package, and both packages rely on the PNAS LaTeX macros. Both these sources are gratefully acknowledged as this work would not have been possible without them. Our extensions are under the same respective licensing term (GPL-3 and LPPL (>= 1.3)).

References

Deurenberg P, Deurenberg-Yap M, Guricci S (2002). "Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship." *Obesity Reviews*, **3**(3), 141–146. doi:10.1046/j.1467-789x.2002.00065.x.

Deurenberg P, Yap M (1999). "The assessment of obesity: Methods for measuring body fat and global prevalence of obesity." *Best Practice and Research Clinical Endocrinology and Metabolism*, **13**(1), 1–11. doi:10.1053/beem.1999.0003.

Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, Mozaffarian D, Swinburn B, Ezzati M (2019). "The obesity transition: Stages of the global epidemic." *The Lancet Diabetes and Endocrinology*, **7**(3), 231–240. doi:10.1016/s2213-8587(19)30026-9.

Penrose KW, Nelson AG, Fisher AG (1985). "Generalized body composition prediction equation for men using simple measurement techniques." *Medicine and Science in Sports and Exercise*, **17**(2), 189. doi:10.1249/00005768-198504000-00037.

Trang LT, Trung NN, Chu DT, Hanh NT (2019). "Percentage body fat is as a good indicator for determining adolescents who are overweight or obese: A cross-sectional study in Vietnam." *Osong Public Health and Research Perspectives*, **10**(2), 108–114. doi:10.24171/j.phrp.2019.10.2.10.