

# Fitting Percentage of Body Fat to Simple Body Measurements

Roger W. Johnson  
Carleton College

*Journal of Statistics Education* v.4, n.1 (1996)

Copyright (c) 1996 by Roger W. Johnson, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

---

**Key Words:** Multiple regression.

## Abstract

Percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) are recorded for 252 men. Body fat, one measure of health, has been accurately estimated by an underwater weighing technique. Fitting body fat to the other measurements using multiple regression provides a convenient way of estimating body fat for men using only a scale and a measuring tape. This dataset can be used to show students the utility of multiple regression and to provide practice in model building.

## 1. Introduction

1 A variety of popular health books suggest that readers assess their health, at least in part, by estimating their percentage of body fat. [Bailey \(1994, pp. 179-186\)](#), for instance, presents tables of estimates based on age, gender, and various skinfold measurements obtained using a caliper. [Bailey \(1991, p. 18\)](#) suggests that "15 percent fat for men and 22 percent fat for women are maximums for good health." [Behnke and Wilmore \(1974, pp. 66-67\)](#), [Wilmore \(1976, p. 247\)](#), [Katch and McArdle \(1977, pp. 120-132\)](#), and [Abdel-Malek, et al. \(1985\)](#) are other sources of predictive equations for body fat. These predictive equations use skinfold measurements, body circumference measurements (e.g., abdominal circumference), and, in the Abdel-Malek article, simply height and weight. [Gardner and Poehlman \(1993, 1994\)](#) supplement these body measurements with a measure of physical activity to predict body density from which, as we shall see below, body fat can be estimated.

2 Such predictive equations for the determination of body fat can be determined through multiple regression. A group of subjects is gathered, and various body measurements and an accurate estimate of the percentage of body fat are recorded for each. Then body fat can be fit to the other measurements using multiple regression, giving, we hope, a useful predictive equation for people similar to the subjects. The various measurements other than body fat recorded on the subjects are, implicitly, ones that are easy to obtain and serve as proxies for body fat, which is not so easily obtained.

3 In the dataset provided by Dr. A. Garth Fisher (personal communication, October 5, 1994), age, weight, height, and 10 body circumference measurements are recorded for 252 men. Each man's percentage of body fat was accurately estimated by an underwater weighing technique discussed below. A complete listing of the variables in the dataset appears in the Appendix.

## 2. Determination of the Percentage of Body Fat from Underwater Weighing

4 The percentage of body fat for an individual can be estimated from body density. As an approximation, assume that the body

consists of two components -- lean tissue and fat tissue. Letting

D = body density,  
W = body weight,  
A = proportion of lean tissue,  
B = proportion of fat tissue (so  $A + B = 1$ ),  
a = density of lean tissue, and  
b = density of fat tissue,

we have

D = weight/volume  
=  $W / [\text{lean tissue volume} + \text{fat tissue volume}]$   
=  $W / [A \cdot W / a + B \cdot W / b]$   
=  $1 / [(A/a) + (B/b)]$ .

Solving for B we find

$$B = (1/D) * [ab/(a - b)] - [b/(a - b)].$$

5 Using the estimates  $a = 1.10 \text{ gm/cm}^3$  and  $b = 0.90 \text{ gm/cm}^3$  (see [Katch and McArdle 1977, p. 111](#), or [Wilmore 1976, p. 123](#)), we come up with "Siri's equation" ([Siri 1956](#)):

$$\text{Percentage of body fat (i.e., } 100 * B) = 495/D - 450,$$

where D is in units of  $\text{gm/cm}^3$ . The dataset provided also gives a second estimate of body fat due to [Brozek, Grande, Anderson, and Keys \(1963, p. 137\)](#):

$$\text{Percentage of body fat} = 457/D - 414.2,$$

which is considered accurate for "individuals in whom the body weight has been free from large, recent fluctuations." There does not seem to be uniform agreement in the literature as to which of these two methods is best.

6 Volume, and hence the body density D, can be accurately measured in a variety of ways. The technique of underwater weighing "computes body volume as the difference between body weight measured in air and weight measured during water submersion. In other words, body volume is equal to the loss of weight in water with the appropriate temperature correction for the water's density" ([Katch and McArdle 1977, p. 113](#)). Using this technique,

$$\text{Body density} = W / [(W - WW) / \text{c.f.} - LV],$$

where

W = weight in air (kg)  
WW = weight in water (kg)  
c.f. = water correction factor  
(equal to 1 at 39.2 degrees F because one gram of  
water occupies exactly one  $\text{cm}^3$  at this temperature,  
equal to .997 at 76-78 degrees F)  
LV = residual lung volume (liters)

([Katch and McArdle 1977, p. 115](#)). The dataset provided here contains the weights of the subjects, but not the values of the three other quantities. Other methods of determining body volume are given in [Behnke and Wilmore \(1974, p. 22 ff.\)](#).

### 3. Student Explorations

7 I have presented this dataset to my students after I have discussed multiple regression and have illustrated, in the lab with another

dataset, some techniques that they might try (e.g., plots of dependent versus independent variables, residual plots, the use of transformations of the independent variables in the model) when trying to build a regression model. They work in pairs on the following questions after I have given them some background on the variables in the dataset.

8 (a) Examine the data and note any unusual cases. Sort the cases, for example, by height, weight, and percentage of fat and note the distributions. What should be done, if anything, about these unusual cases? Suggest some rules for changing or deleting outliers.

Comments: Much to the dismay of some students, there are a few apparent errors in the dataset. Case 42, for instance, apparently weighs 205 pounds, but measures only 29.5 inches in height! Fortunately, we can infer the correct values from other variables in the file. The lean body weight or fat free weight of this individual is listed as 140.1 pounds, which is, up to rounding,  $(1 - \text{fraction of body fat using Brozek's equation}) * 205$ . Consequently, the listed weight is probably correct. From the adiposity index of 29.9 kg/meters<sup>2</sup>, which is weight divided by height<sup>2</sup>, one can infer that the height should probably be 69.5 inches instead of 29.5 inches (a change in just one of the digits). One can check for internal consistency between other variables as well. In cases 48, 76, and 96, for instance, the density values do not give rise to the two estimates of body fat percentage recorded. In each case, a change of a single digit in the density gives the body fat percentages indicated for that individual. In particular, it seems that the following changes (among others) are in order:

Case	Listed Body Density	Apparently Correct Body Density
----	-----	-----
48	1.0665	1.0865
76	1.0666	1.0566
96	1.0991	1.0591

Such errors help students become more aware of data integrity issues. Also note that case 182 is a particularly lean individual whose predicted percentage of body fat is negative according to Siri's and Brozek's equation and has been truncated to zero in the dataset.

9 (b) Choose one of the two percentage of body fat estimates, either Brozek's method or Siri's method. Fit this percentage of body fat in terms of some subset of the remaining variables excluding density, which is not easily measured. The researchers who collected these data, [Penrose, Nelson, and Fisher \(1985\)](#), built a regression model for fat free weight =  $(1 - \text{fraction of body fat}) * \text{weight}$  that used the variables weight, age, age<sup>2</sup>, height, and (abdomen - wrist). Do you find any of these variables useful in fitting percentage of body fat? September 14, 1995 articles in *The New England Journal of Medicine* link high values of the adiposity index (weight/height<sup>2</sup>), sometimes called the body mass index, to increased risk of premature death. See if this variable is useful in your model. Also try weight<sup>1.2</sup>/height<sup>3.3</sup> as suggested in [Abdel-Malek, et al. \(1985\)](#). Why should one bother to fit percentage of body fat using these other variables?

10 (c) (This question requires a tape measure and/or a scale.) Estimate the percentage of body fat for each member of your group using your regression model. Is this model appropriate for all the members of your group? How about for other people in class? What is the most general audience to which your model can be applied? (Note: 1 inch = 2.54 cm.)

Comment: Most students see that the model should not be used for women; other students get into more subtle issues regarding the age of college students compared to the age of the men in the dataset.

11 (d) Comment on the accuracy of your model. Discuss, in particular, what the standard error means. What kind of error should a user of this model expect as opposed to the prediction error for those folks who were used to build the model?

Comment: Before starting this lab I have already discussed the "incestuous" nature of the standard error value and how the model is likely to give an error larger than the standard error for cases that were not used to build it.

12 (e) Estimate the percentage of American men with percentage of body fat less than 15% (the maximum for good health given by [Bailey \(1991\)](#) above). What assumptions did you make?

13 (f) (Advanced) [Penrose, Nelson, and Fisher \(1985\)](#) built their regression model using just the first 143 cases of the 252 cases in the dataset. The remaining 109 cases were used to get a true estimate of the error of the model (c.f. question d). Here is an alternative "cross-validation" approach to error estimation I would like you to consider (but do not actually perform):

- i. Build a regression model using all 252 cases.
- ii. Using the model (i.e., variables) above, refit the model coefficients using all the cases but the first and record the error in body fat using this model on the first case.
- iii. Repeat ii., leaving out, in turn, just the second, just the third, ... , just the 252nd case, each time recording the error.
- iv. Compute the standard deviation of the 252 errors to provide an estimate of the accuracy of the model.

14 Why is the resulting error estimate (the standard deviation in iv.) a better estimate of the true error that one should expect in using the model than the standard error? Discuss the advantages and disadvantages of this alternative cross-validation procedure compared to what Penrose, Nelson, and Fisher did.

15 In keeping with the other data analysis labs undertaken during the term, I serve as a resource/coach for student pairs during the class period in which they initially look at these data. After this class period the student pairs arrange to meet outside of class to continue their analysis and eventually write up their findings in a polished, word-processed report. Further details may be found in [Egge, Foley, Haskins, and Johnson \(1994, 1995\)](#).

## 4. Concluding Notes

16 My students have enjoyed working with this dataset and, upon hearing that I have a caliper and tables to estimate body fat from various skinfold measurements, come to me to borrow them! I show the students a few of the regression models produced after the assignment is due to remind them how subjective the model building process is and how different people can come up with some rather different, but perhaps equally effective, models.

## 5. Getting the Data

17 The file [fat.dat](#) contains the raw data. The file [fat.txt](#) is a documentation file containing a brief description of the dataset.

## Acknowledgments

Thanks to Dr. A. Garth Fisher (1994) who generously provided the dataset to freely distribute and use for non-commercial purposes. Thanks also to Richard Wetzel, M.D., for enlightening correspondence about some of the difficulties involved in body fat estimation and for tracking down some references.

---

## Appendix - Key to Variables in fat.dat

### Columns

- 3 - 5 Case Number
- 10 - 13 Percent body fat using Brozek's equation,  
457/Density - 414.2
- 18 - 21 Percent body fat using Siri's equation,  
495/Density - 450
- 24 - 29 Density (gm/cm<sup>3</sup>)
- 36 - 37 Age (yrs)

40 - 45 Weight (lbs)  
 49 - 53 Height (inches)  
 58 - 61 Adiposity index =  $\text{Weight}/\text{Height}^2$  (kg/m<sup>2</sup>)  
 65 - 69 Fat Free Weight  
     = (1 - fraction of body fat) \* Weight,  
     using Brozek's formula (lbs)  
 74 - 77 Neck circumference (cm)  
 81 - 85 Chest circumference (cm)  
 89 - 93 Abdomen circumference (cm) "at the umbilicus  
     and level with the iliac crest"  
 97 - 101 Hip circumference (cm)  
 106 - 109 Thigh circumference (cm)  
 114 - 117 Knee circumference (cm)  
 122 - 125 Ankle circumference (cm)  
 130 - 133 Extended biceps circumference (cm)  
 138 - 141 Forearm circumference (cm)  
 146 - 149 Wrist circumference (cm) "distal to the  
     styloid processes"

---

## References

- Abdel-Malek, A. K., Mukherjee, D., and Roche, A. F. (1985), "A Method of Constructing an Index of Obesity," *Human Biology*, 57(3), 415-430.
- Bailey, C. (1991), *The New Fit or Fat*, Boston: Houghton-Mifflin.
- (1994), *Smart Exercise: Burning Fat, Getting Fit*, Boston: Houghton-Mifflin.
- Behnke, A., and Wilmore, J. (1974), *Evaluation and Regulation of Body Build and Composition*, Englewood Cliffs, N.J.: Prentice Hall.
- Brozek, J., Grande, F., Anderson, J., and Keys, A. (1963), "Densitometric Analysis of Body Composition: Revision of Some Quantitative Assumptions," *Annals of the New York Academy of Sciences*, 110, 113-140.
- EGGE, E., FOLEY, S., HASKINS, L., and JOHNSON, R. (1994), "A Data Analysis Based Elementary Statistics Course," in *Proceedings of the Section on Statistical Education, American Statistical Association*, pp. 144-149.
- (1995), *Statistics Lab Manual* (3rd ed.), Department of Mathematics and Computer Science, Carleton College.
- Gardner, A. W., and Poehlman, E. T. (1993), "Physical Activity Is a Significant Predictor of Body Density in Women," *American Journal of Clinical Nutrition*, 57, 8-14.
- (1994), "Leisure Time Physical Activity Is a Significant Predictor of Body Density in Men," *Journal of Clinical Epidemiology*, 47(3), 283-291.
- Katch, F., and McArdle, W. (1977), *Nutrition, Weight Control, and Exercise*, Boston: Houghton Mifflin.
- Penrose, K., Nelson, A., and Fisher, A. (1985), "Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques" (abstract), *Medicine and Science in Sports and Exercise*, 17(2), 189.
- Siri, W. E. (1956), "Gross Composition of the Body," in *Advances in Biological and Medical Physics* (Vol. IV), eds. J. H. Lawrence and C. A. Tobias, New York: Academic Press.

Wilmore, J. (1976), *Athletic Training and Physical Fitness: Physiological Principles of the Conditioning Process*, Boston: Allyn and Bacon.

---

Roger W. Johnson  
Department of Mathematics and Computer Science  
Carleton College  
Northfield, MN 55057-4001  
[rjohnson@carleton.edu](mailto:rjohnson@carleton.edu)

---

[Return to Table of Contents](#) | [Return to the JSE Home Page](#)