**Context**

Lists estimates of the percentage of body fat determined by underwater  
weighing and various body circumference measurements for 252 men.

**Educational use of the dataset**

This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly.

**Content**

The variables listed below, from left to right, are:

1. Density determined from underwater weighing
2. Percent body fat from Siri's (1956) equation
3. Age (years)
4. Weight (lbs)
5. Height (inches)
6. Neck circumference (cm)
7. Chest circumference (cm)
8. Abdomen 2 circumference (cm)
9. Hip circumference (cm)
10. Thigh circumference (cm)
11. Knee circumference (cm)
12. Ankle circumference (cm)
13. Biceps (extended) circumference (cm)
14. Forearm circumference (cm)
15. Wrist circumference (cm)

(Measurement standards are apparently those listed in Benhke and Wilmore (1974), pp. 45-48 where, for instance, the abdomen 2 circumference is measured "laterally, at the level of the iliac crests, and anteriorly, at the umbilicus".)

These data are used to produce the predictive equations for lean body weight given in the abstract "Generalized body composition prediction equation for men using simple measurement techniques", K.W. Penrose, A.G. Nelson, A.G. Fisher, FACSM, Human Performance Research Center, Brigham Young University, Provo, Utah 84602 as listed in *Medicine and Science in Sports and Exercise*, vol. 17, no. 2, April 1985, p. 189. (The predictive equation were obtained from the first 143 of the 252 cases that are listed below).

**More details**

A variety of popular health books suggest that the readers assess their health, at least in part, by estimating their percentage of body fat. In Bailey (1994), for instance, the reader can estimate body fat from tables using their age and various skin-fold measurements obtained by using a caliper. Other texts give predictive equations for body fat using body circumference measurements (e.g. abdominal circumference) and/or skin-fold measurements. See, for instance, Behnke and Wilmore (1974), pp. 66-67; Wilmore (1976), p. 247; or Katch and McArdle (1977), pp. 120-132).

The percentage of body fat for an individual can be estimated once body density has been determined. Folks (e.g. Siri (1956)) assume that the body consists  
of two components - lean body tissue and fat tissue. Letting:

* D = Body Density (gm/cm^3)
* A = proportion of lean body tissue
* B = proportion of fat tissue (A+B=1)
* a = density of lean body tissue (gm/cm^3)
* b = density of fat tissue (gm/cm^3)

we have:

D = 1/[(A/a) + (B/b)]

solving for B we find:

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

Using the estimates a=1.10 gm/cm^3 and b=0.90 gm/cm^3 (see Katch and McArdle (1977), p. 111 or Wilmore (1976), p. 123) we come up with "Siri's equation":

Percentage of Body Fat (i.e. 100\*B) = 495/D - 450.

Volume, and hence body density, can be accurately measured 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,

Body Density = WA/[(WA-WW)/c.f. - LV]

where:

* WA = Weight in air (kg)
* WW = Weight in water (kg)
* c.f. = Water correction factor (=1 at 39.2 deg F as one-gram of water occupies exactly one cm^3 at this temperature, =.997 at 76-78 deg F)
* LV = Residual Lung Volume (liters)

(Katch and McArdle (1977), p. 115). Other methods of determining body volume are given in Behnke and Wilmore (1974), p. 22 ff.

**Source**

The data were generously supplied by Dr. A. Garth Fisher who gave permission to freely distribute the data and use for non-commercial purposes.

Roger W. Johnson  
Department of Mathematics & Computer Science  
South Dakota School of Mines & Technology  
501 East St. Joseph Street  
Rapid City, SD 57701

email address: [rwjohnso@silver.sdsmt.edu](mailto:rwjohnso@silver.sdsmt.edu)  
web address: <http://silver.sdsmt.edu/~rwjohnso>

**References**

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

Behnke, A.R. and Wilmore, J.H. (1974). *Evaluation and Regulation of Body Build and Composition*, Prentice-Hall, Englewood Cliffs, N.J.

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.

Katch, Frank and McArdle, William (1977). *Nutrition, Weight Control, and Exercise*, Houghton Mifflin Co., Boston.

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

**Dataset Description: Body Fat Percentage and Circumference Measurements**

This dataset consists of **252 male subjects**, with estimates of **body fat percentage** obtained through **underwater weighing** and various **body circumference measurements**. The primary goal is to develop **multiple regression models** to predict body fat percentage using easy-to-measure body parameters, avoiding the inconvenience and cost of traditional methods like underwater weighing.

The dataset is widely used in **health assessment**, **body composition studies**, and **predictive modeling** in **sports science and medicine**. The equations derived from this dataset are useful for estimating **lean body mass** and **fat percentage** based on simple measurements.

**Table: Dataset Attributes and Description**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| **Density** | Body density determined by underwater weighing (gm/cm³) |
| **Body Fat %** | Percent body fat estimated using Siri’s equation |
| **Age** | Age of the subject (years) |
| **Weight** | Body weight (lbs) |
| **Height** | Height (inches) |
| **Neck Circumference** | Neck circumference (cm) |
| **Chest Circumference** | Chest circumference (cm) |
| **Abdomen Circumference** | Abdomen circumference at iliac crest (cm) |
| **Hip Circumference** | Hip circumference (cm) |
| **Thigh Circumference** | Thigh circumference (cm) |
| **Knee Circumference** | Knee circumference (cm) |
| **Ankle Circumference** | Ankle circumference (cm) |
| **Biceps Circumference** | Biceps (extended) circumference (cm) |
| **Forearm Circumference** | Forearm circumference (cm) |
| **Wrist Circumference** | Wrist circumference (cm) |

**Application in Research & Healthcare**

* **Predicting body fat percentage** using circumference-based models
* **Alternative to expensive & inconvenient fat measurement techniques**
* **Regression modeling for body composition analysis**
* **Health risk assessments based on body fat estimation**

### \*\*Short Analysis of Correlation with Body Fat (%)\*\*

The correlation matrix reveals the relationships between body fat percentage and various body measurements. The key observations are:

1. \*\*Strongest Positive Correlations with Body Fat:\*\*

- \*\*Abdomen (0.813)\*\*: The strongest predictor of body fat. A larger abdominal circumference strongly correlates with higher body fat.

- \*\*Chest (0.703)\*\* and \*\*Hip (0.625)\*\*: These measurements also show a strong correlation with body fat, suggesting that fat distribution in these areas is highly relevant.

- \*\*Weight (0.612)\*\*: Higher body weight is associated with higher body fat but is not the strongest predictor since muscle mass also contributes to weight.

- \*\*Neck (0.491)\*\* and \*\*Biceps (0.493)\*\*: Moderate correlation, indicating that even upper-body measurements contribute to fat estimation.

2. \*\*Strongest Negative Correlations with Body Fat:\*\*

- \*\*Density (-0.988)\*\*: As expected, body density has a nearly perfect inverse relationship with body fat. This aligns with Siri’s equation, where density is used to calculate fat percentage.

- \*\*Height (-0.089)\*\*: Almost negligible correlation, meaning height alone does not predict body fat effectively.

### \*\*Most Influential Predictor:\*\*

- \*\*Abdomen circumference\*\* is the most influential factor (\*\*0.813 correlation\*\*) in predicting body fat. This aligns with medical research showing that central adiposity (fat around the abdomen) is a strong indicator of overall body fat percentage.

### \*\*Conclusion:\*\*

For body fat estimation, abdominal circumference is the most important predictor, followed by chest, hip, and weight. Density remains a theoretical measure rather than a practical input for ML models.

