

## Stats 4590 Project: Body Fat Data

In this project, you will analyze a data set on the percentage of body fat (a measure of health) for a sample of 252 men. Body fat is estimated through an underwater weighing technique. However, your medical colleagues would like to come up with a way to estimate body fat for men using a simpler method, based on easy to obtain body measurements and information. The variables that are available to you are described below.

The data is on the course website, and can be loaded into R using the `read.table` command.

Your goal is to develop a model that estimates the percentage of body fat from some of the available explanatory variables (Are there some variables that it would be inappropriate for you to use in your model?). Since there are 2 measures of body fat provided, you can work with either of the 2 measures in developing your model (**do not use both**). As always, you want to develop a model that is as simple as possible, yet has reasonable predictive skill.

In developing your model, watch out for any unusual cases that may exist in the data.

The columns of the dataset contain the following variables. Each row corresponds to a different subject:

1. Case Number
2. Percent body fat using Brozek's equation,  $457/Density - 414.2$
3. Percent body fat using Siri's equation,  $495/Density - 450$
4. Density ( $gm/cm^3$ )
5. Age (yrs)
6. Weight (lbs)
7. Height (inches)
8. Adiposity index =  $Weight/Height^2$  ( $kg/m^2$ ). This is often called the body mass index (BMI).
9. Fat Free Weight =  $(1 - \text{fraction of body fat}) \times \text{Weight}$ , using Brozek's formula (lbs)
10. Neck circumference (cm)
11. Chest circumference (cm)
12. Abdomen circumference (cm) "at the umbilicus and level with the iliac crest"
13. Hip circumference (cm)
14. Thigh circumference (cm)
15. Knee circumference (cm)
16. Ankle circumference (cm)
17. Extended biceps circumference (cm)
18. Forearm circumference (cm)
19. Wrist circumference (cm) "distal to the styloid processes"