

For the “Introduction to the Advanced Lab” tutorial you explored data collected on athletes at the Australian Institute of Sport. This assignment uses the following six variables from that data set:

<b>Sex</b>	1=female	2=male
<b>BMI</b>	body mass index (weight/height <sup>2</sup> )	
<b>SSF</b>	sum of skinfolds	
<b>Bfat</b>	percentage bodyfat	
<b>Ht</b>	height (m)	
<b>Wt</b>	weight (kg)	

In this assignment you are asked to explore some aspects of this dataset and to write a report that clearly communicates your findings.

Percent body fat (Bfat) is a key element of overall fitness. Unfortunately, it is difficult to measure directly. One direct method is called “Hydrodensitometry” or “Underwater Weighing”. This method measures whole body density by determining body volume. First a person is weighed in the usual manner and then weighed again when total immersed in water. The densities of bone and muscles are higher than water, and fat is less dense than water. So a person with more bone and muscle will weigh more in water than a person with less bone and muscle, meaning they have a higher body density and lower percentage of body fat.

An alternative approach is to create a regression model that predicts Bfat using variables that are easier to measure. One variable that is often used is skinfold thickness (SSF). Hand-held callipers that exert a standard pressure are used to measure the skinfold thickness at various body locations (3-7 test sites are common). Then a calculation is used to estimate the body fat percentage for a person based on the sum of these measurements.

1. Using the Australian Institute of Sport data, create a simple regression model that predicts Bfat using just SSF. Evaluate the precision of predictions made with this model.
2. Try fitting separate models for the female and male subjects. Does this improve the predictions of Bfat from SSF? Present evidence and at least one plot that supports your answer. Note that using

```
data = sport.df[sport.df$Sex == 1, ]
```

in Splus functions will extract the observations for female athletes and replacing 1 with 2 will do the same for male athletes.

3. The data set contains three additional variables that are very easy to measure: height (Ht), weight (Wt), and body mass index (BMI). Are better predictions of Bfat obtained if one or more of these variables are added to the model? (Note: if you decided that separate models for females and males are preferable try adding the additional variables to each model.) Give reasons to justify your answer.

Your assignment should consist of a concise report (paragraphs, sentences, punctuation marks) and a statistical appendix. You will be given a handout called “Writing up Assignments” that will describe what is expected.

This assignment should be handed in to the appropriate box in the basement of the Maths/Physics building by the SMIS Resource Centre, by 4pm on Friday, 10 August.