

Department of Statistics

COURSE STATS 330/762

Assignment 3, 2010

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm September 16th**

The data set for this assignment is in the file **pbf.txt** which is available on the course web page.

In a study conducted at the Human Performance Research Center, Brigham Young University, percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) were recorded for 252 men. Percentage body fat, a measure of health, is estimated through 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. In actual fact, the percentage of body fat in the data set is a derived quantity, and is calculated from the body density D where D is in units of gm/cm^3 . The relationship is

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

In turn, the density is the ratio of the body weight in grams and the body volume in cubic centimeters. Thus, if we know the weight and volume, we can calculate the percentage of body fat. The weight of an individual can be simply measured with an accurate set of scales. The volume is more of a problem as it involves weighing an individual under water.

It is of interest to develop a statistical model that can predict an individual's volume (and hence their percentage of body fat) from more easily obtained measurements, such as measuring the individual's height and abdomen circumference. These measurements can be made with a tape measure.

To develop such a model, the data supplied can be used. Note that the data has no header row, but the variables in the data set from left to right are

PBF:	Percent body fat, equal to $457/\text{Density} - 414.2$
Density:	Density (gm/cm^3)
Age:	Age (yrs)
Weight:	Weight (lbs)
Height:	Height (inches)
BMI:	Adiposity index = $\text{Weight}/\text{Height}^2$ (kg/m^2)

Neck:	Neck circumference (cm)
Chest:	Chest circumference (cm)
Abdomen:	Abdomen circumference (cm) at the umbilicus and level with the iliac crest
Hip:	Hip circumference (cm)
Thigh:	Thigh circumference (cm)
Knee:	Knee circumference (cm)
Ankle:	Ankle circumference (cm)
Biceps:	Extended biceps circumference (cm)
Forearm:	Forearm circumference (cm)
Wrist:	Wrist circumference (cm) distal to the styloid processes.

1. Read the data into R. Make a data frame, naming the variables with the names above. Print out the first 10 lines. [5 marks]
2. I have not changed any values in the original data set, but there are several strange values. Identify these using graphical methods and delete the offending observations. (Delete a maximum of 4). In particular, some of the PBF values seem suspect (which ones?) Calculate the volume in liters from the variables Density and Weight, and add this variable to the data frame. [5 marks] NB: 1 pound = 453.59237 grams
3. Develop a model that will predict the volume (in litres) from the other variables, excluding Density and PBF. You should be able to come up with a model that predicts very well. Points to note: Which variables should be selected? Are transformations indicated? (think Cherry trees). You should potentially consider using all the techniques you have been taught, up to the end of lecture 15. [20 marks]
4. I have replaced the values of the variables PBF and Density on the last two individuals in the data set with NA's . Using your model, predict the body volume for these two individuals. [10 marks]

NB: There will be a prize for the best predictions. In the event of a tie, a stochastic mechanism will be used.

Extra question for 762 students

Suppose we logged the volume and the other variables (excluding PBF and Density), and fitted a model to log volume, using the other logged variables. Can you explain why we would not need to include the variable $\log(\text{BMI})$ in the model, given the other variables are included?

NB: I have not included a printout of the body fat data: you may assume that all the values in the file are correct.