Stat 628 Module 2: BodyFat

YUKUN FANG   MENGKUN CHEN   JIAYI SHEN

1 Introduction

A variety of popular health books suggest that the readers assess their health, at least in part, by estimating their percentage of body fat. The goal of the project is to come up with a simple and accurate way of deteriming body fat percentage of males based on readily available clinical measurements and find the rule of thumb.
Our rule of thumb:

$$\mathrm{BodyFat} =$$

2 Background

The data set contains measurements from 252 men who had their body fat percentage accurately measured via underwater weighing. Notice that the outcome variable is body fat percentage and he set of predictors are every variable except ID number, body fat percentage, and density.

3 Motivation and Statement

3.1 Motivation

In the beginning we have used linear model step selection by using AIC or BIC to find the best model which may lead to multicollinearity during fitting the model. So we use the principal component analysis to find the PC(principal components), and then use PCs to find linear model.

3.2 Statement of the Model

First, we calculate the covariance matrix of the data set for 15 variables without the three. Then we calculate the eigenvalues and the eigenvectors of the matrix. Then calculate the cumsum of the eigenvalues and divided by sum of eigenvectors and equals to cumvariance. This value is the main result which is use to find the principle components.
Second, we plot the cumvariance and select the top two eigenvalues as the PCs: PC1 and PC2. Then we need to remove the variables whose coefficients are so small with big p-value for t-statistics.
At last, we use PCs as variables to fit the linear model to diagnose and remedy it.

4 Estimation and Inference of Parameters

After the calculation, we get the 2 PCs:

$$PC_1 \sim 0.88 \times \mathrm{WEIGHT} + 0.24 \times \mathrm{CHEST} + 0.30 \times \mathrm{ABDOMEN} + 0.20 \times \mathrm{HIP} + 0.14 \times \mathrm{THIGH}$$

$$PC_2 \sim 0.96 \times \mathrm{AGE}$$