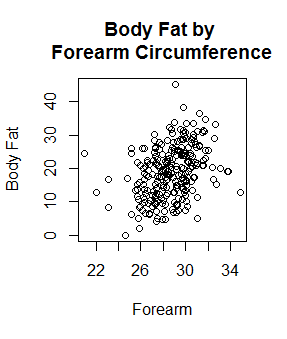
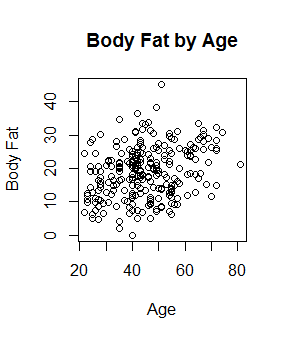
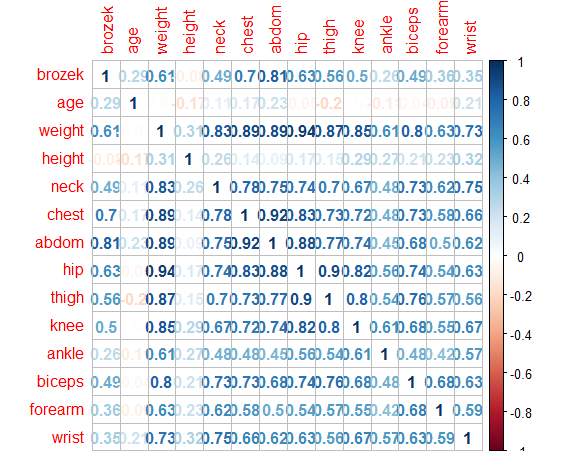
1. The current techniques for measuring body fat are complex and hard to accomplish. One method requires submerging the body in water and measuring the displacement. Researchers are seeking to obtain a simpler way to measure body fat. We have a data set which contains the age, height, weight, and 10 body circumference measurements for 252 men. We will use this data set to create a mathematical model that will hopefully allow us to predict body fat percentage using these measurements.

2. Exploring the data, we automatically see the fact that there are multiple predictor variables. A pairs plot shows that each of these predictor variables maintains a linear relationship with the body fat measurement. We do not show the whole pairs plot here, but below are examples of the body fat measurement plotted against three of the predictor variables. You can see that each has a linear relationship.



We also can create a correlation plot to show the correlation coefficients in relation to every relationship in the data set. This helps us to see which predictor variables have high correlation with body fat measurement, as well as which predictor variables have high correlation with each other.



A multiple linear regression model is appropriate for these data.

3.

yi = β0 + β1i) + β2i ) + β3i) + β4i ) + β5i) + β6i) + β7i) + β8i) + β9i) + β10i) + β11i) + β12i) + β13i) + εi

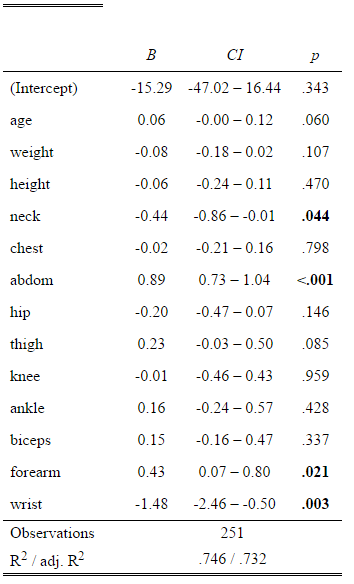
where εi ~ N(0,σ2)

yi = the body fat measurement for the ith person

β0 = If you center each of your x’s, then β0 is the value of the body fat measurement when all the x’s are at their average and slopes don’t change.

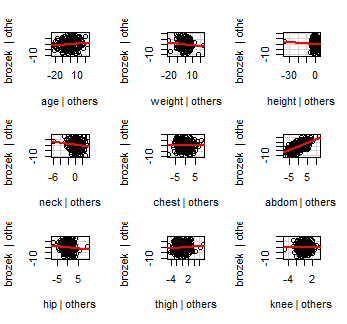
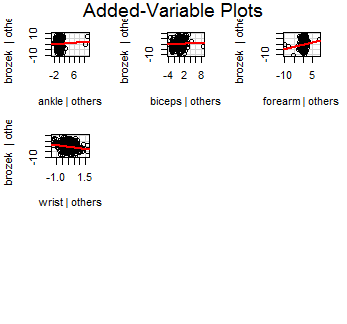
β1 = Holding all other x's constant, as you increase age by 1, β1 will be the average increase in body fat.

This model assumes linearity, independence, normality, and equal variance

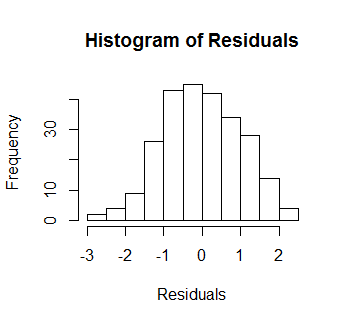
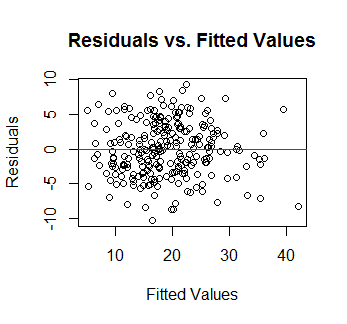
4.

The table to the left contains the results of fitting our model to the data. It shows the β0, as explained in the previous question, as well as the coefficients for all of the other variables. Let's take the abdomen for example. As the circumference of the abdomen goes up by 1, we can expect the body fat measurement to increase by .89, on average.

5. We must show that our assumptions are satisfied in order for the MLR model to be justifiable. We already explored linearity earlier, but we can explore in more depth by looking at the Added-Variable plots. These plots show a regression line fitted for each individual predictor variable and the body fat measurement. We can see linearity in all of these relationships.



Looking at the residuals versus predicted values below, the data appears to have equal variance across the line. There are also no apparent patterns in this plot. The assumptions for equal variance and independence are satisfied. The Histogram of Residuals shows us that the errors are normally distributed. It doesn't even look like there are any outliers. The assumption of normality is also satisfied.



6. With the age, height, weight, and list of ten circumference measurements, we can just plug those values into our model and it will give us a predicted body fat measurement. In the case where: age= 50, weight= 203, height= 67, neck= 40.2, chest=114.8, abdom=108.1, , hip=102.5, thigh=61.3, knee= 41.1, ankle= 24.7, biceps= 34.1, forearm= 31, and wrist= 18.3, we obtain a body fat prediction of 31.3059. We are 95% confident that the true body fat measurement lies between 23.20 and 39.42.

7. We ran a cross-validation simulation in order to assess the accuracy of our predictions. We found a bias of .0017, almost 0. This means that are predictions, on average, are centered correctly and are not too high, nor too low. We also found a RPMSE of 3.833, which means that our predictions are off by about 3.833 on average. The prediction intervals that are generated with the prediction are covering the true body fat measurement 97.27% of the time. Those intervals have an average width of 16.17. These statistics show that our intervals are very accurate, but they are also fairly wide, making them less useful the wider they get.