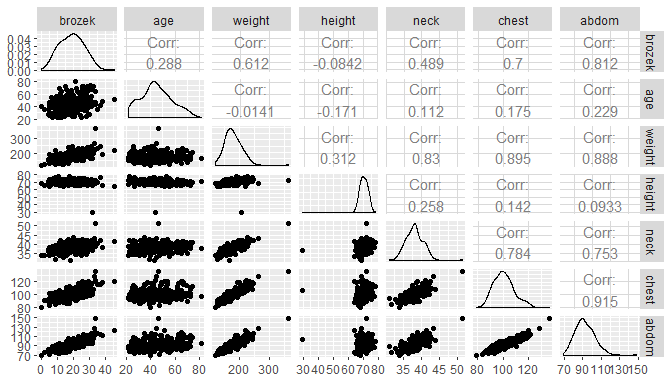
HW-4

Chris Trippe & Ethan Kemeny

1. There are many ways to measure body fat, but researchers are seeking out a simpler method. One technique is submerging a body in a tank of water and measuring the increase in water level. The submersion method was used to measure the percentage of fat of 252 men, then their age, weight, heigh and 10 body circumference measurements were taken. The given data will be used to see what variables can be used to best measure a person’s body fat percentage. A model can be created from this data to better predict body fat percentage without going through more complex processes such as full body water submersion.
2.  This plot shows several of the variables and their correlation values. We are mostly interested in how the variables relate to body fat percentage. We can see that they look fairly linear. That is, as the variables increase, so does body fat percentage for all variables except height. Looking at the correlation values we can see that for body fat percentage they are all positive except for height once again. We will pull those out to get a better look.

age weight height neck chest abdom  
[1,] 0.2881482 0.6118683 -0.08417553 0.488992 0.6997151 0.8120161  
 hip thigh knee ankle biceps forearm wrist  
[1,] 0.6267491 0.5619947 0.5048113 0.2625731 0.4920415 0.3586868 0.348876

Here we can see the correlation between the variable and body fat percentage. Most of them are fairly strong except age, height, ankle, forearm, and wrist. With height, there seems to be almost no correlation.

is the person’s body fat percentage.

is the person’s explanitory variable measurement. For example, ith the person’s age.

is the y-intercept; on average, what you would expect a person’s body fat percentage to be when all explanatory variables are equal to zero.

is the slope coefficient that pertains to age; as a person’s age increases by 1 year, a person’s body fat is expected to increase by percent, on average, assuming all other explanatory variables are held constant.

is the slope coefficient that pertains to weight; as a person’s weight increases by 1 pound, a person’s body fat is expected to increase by percent, on average, assuming all other explanatory variables are held constant.

is the slope coefficient that pertains to wrist circumference; as a person’s wrist circumference increases by 1 centimeter, a person’s body fat is expected to increase by percent, on average, assuming all other explanatory variables are held constant.

is the residual; the error associated with the person’s explanitory variable.

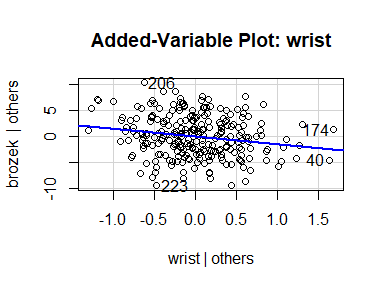
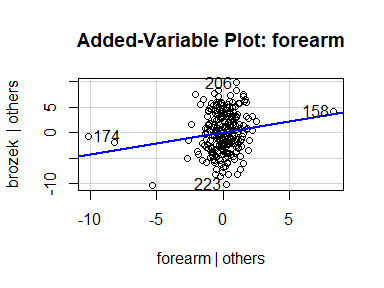
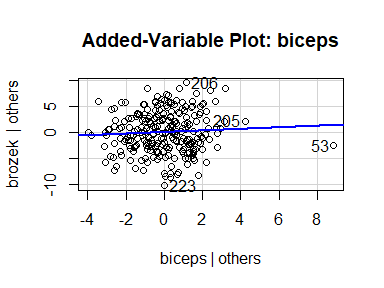
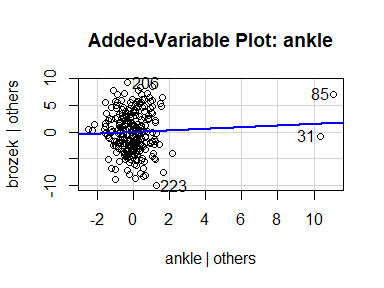
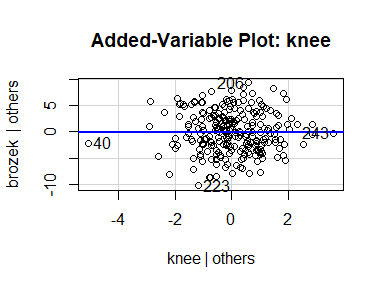
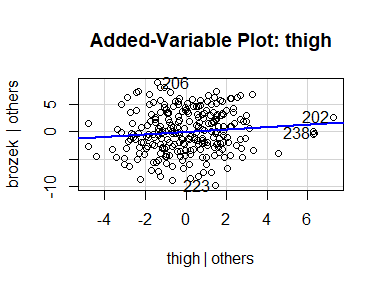
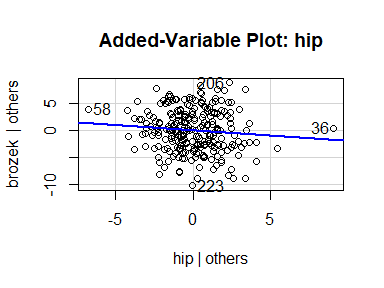
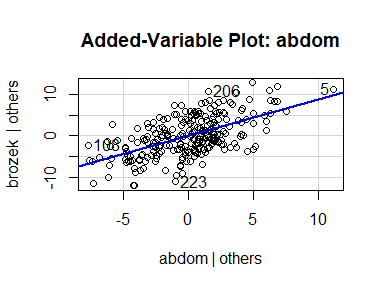
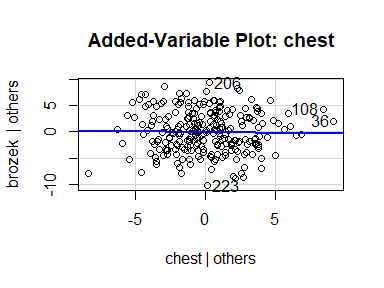
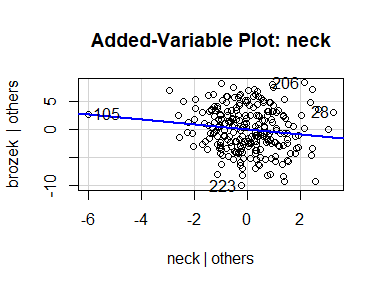
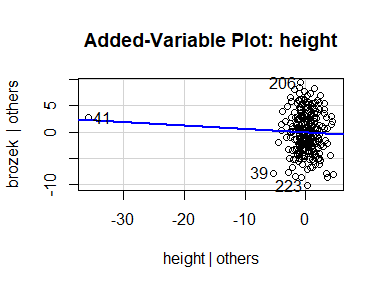
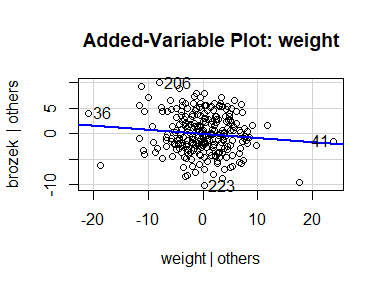
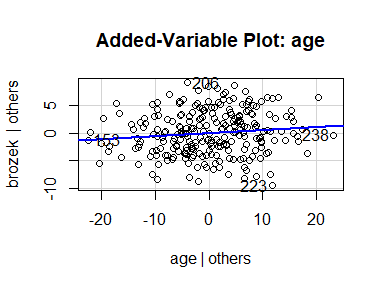
On average, we expect a person’s body fat percentage to be -15.2923 when all explanatory variables are equal to zero.

As a person’s age increases by 1 year, a person’s body fat is expected to increase by 0.0568 percent, on average, assuming all other explanatory variables are held constant.

As a person’s weight increases by 1 pound, a person’s body fat is expected to decrease by 0.0803 percent, on average, assuming all other explanatory variables are held constant.

As a person’s wrist circumference increases by 1 centimeter, a person’s body fat is expected to decrease by 1.4766 percent, on average, assuming all other explanatory variables are held constant.

In order to use a multiple linear regression model, the model must be linear, a person’s body fat percentage must be independent of all other workers body fat percentage, the residuals must be normally distributed about the fitted line, and the residuals must have equal variance about the line.



They all look approximately linear, though some relationships may be weak, such as height, ankle, and wrist.

studentized Breusch-Pagan test  
  
data: body\_slr  
BP = 19.072, df = 13, p-value = 0.1209

The p-value for the Breusch-Pagan test is 0.1209 which is greater than 0.5. This means we fail to reject the null hypothesis and conclude that the residuals have equal variance about the line.

One-sample Kolmogorov-Smirnov test  
  
data: stdres(body\_slr)  
D = 0.039776, p-value = 0.822  
alternative hypothesis: two-sided

The p-value for the Kolmogorov-Smirnov test is 0.822 which is greater than 0.5. This means we fail to reject the null hypothesis and conclude that the residuals are normally distributed about the line.

It’s safe to assume that the body fat percentages of each person are independent. Some of the only worries would be that if all these people come from the same company, their activity levels may be the same as well as what they eat if they have a cafeteria or certain restaurants nearby. However, this will be considered negligent and we’ll conclude that the body fat percentages are independent.

Using a cross-validation study of 500 samples we are able to get a sense of how accurate our model is at predicting body fat percentage. The bias is 0.0273. This means our model typically predicts body fat percentage 0.0273 above the actual value on average. That value is relatively small so we feel our model isn’t very biased. The root prediction mean squared error value of 0.7027 tells us our predictions miss the mark by an average of 0.7027.

We also wanted to look at our prediction intervals or the low and high end of our estimated values. The study returned a coverage of 0.95144 which is the percentage of prediction intervals that contain the actual value. With a width of 16.3033391, we can conclude that our prediction intervals are about 16.3033391% wide which is fairly large which would make sense of why our coverage is also large.

Our model has an value of 0.7464. This means that the variables together are able to explain 74.64% of the variation in body fat percentage. This means our model does a good job of fitting the data. It is not the best because there is still a variation on body fat we can’t explain, but it is enough to feel comfortable using our model.

This value with the other statistics explained above shows that our model does a fairly good job at fitting the data and predicting body fat percentage.

#7 For this person we would enter all of the values of age, weight, height, etc. into our linear model. Then using the coefficents from the model we would calculate the estimated body fat percentage.

Using the model we get a prediction of 31.3035886. This means for a person with the measurements given, we would expect them to have 31.3035886% body fat. We are 95% confident that the person’s actual body fat percentage would be between 23.1919228% and 39.4152545%.