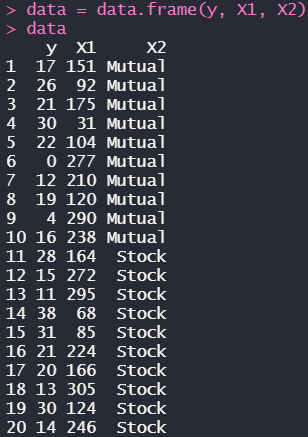
**STAT 40001/STAT 59800 Statistical Computing Fall 2020**

**Lab -19**

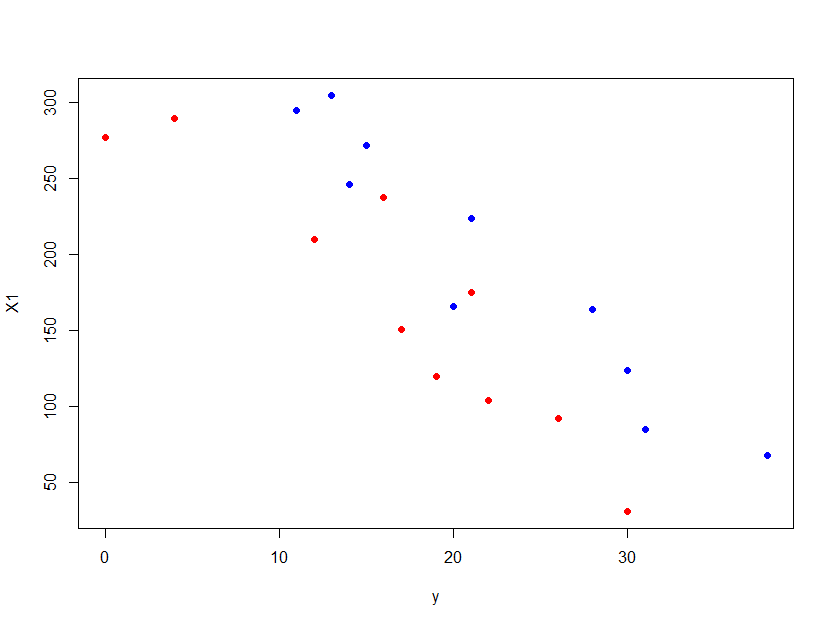
1. An economist studied 10 mutual firms and 10 stock firms. Let y= Number of months elapsed, X1= Size of the firm and x2= Type of firm. Below is the data

|  |  |  |
| --- | --- | --- |
| y | X1 | X2 |
| 17 | 151 | Mutual |
| 26 | 92 | Mutual |
| 21 | 175 | Mutual |
| 30 | 31 | Mutual |
| 22 | 104 | Mutual |
| 0 | 277 | Mutual |
| 12 | 210 | Mutual |
| 19 | 120 | Mutual |
| 4 | 290 | Mutual |
| 16 | 238 | Mutual |
| 28 | 164 | Stock |
| 15 | 272 | Stock |
| 11 | 295 | Stock |
| 38 | 68 | Stock |
| 31 | 85 | Stock |
| 21 | 224 | Stock |
| 20 | 166 | Stock |
| 13 | 305 | Stock |
| 30 | 124 | Stock |
| 14 | 246 | Stock |

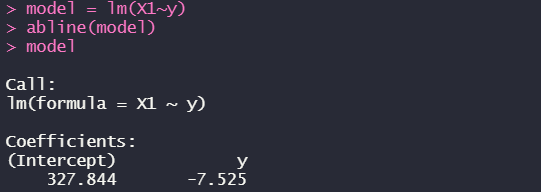
1. Draw a scatter plot of Size of the firm vs. Number of month elapsed. Also choose different colors to display Type of the firm.





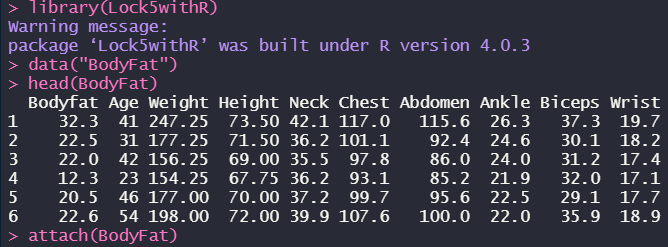


1. Fit a regression model with indicator variable and write out the regression model.



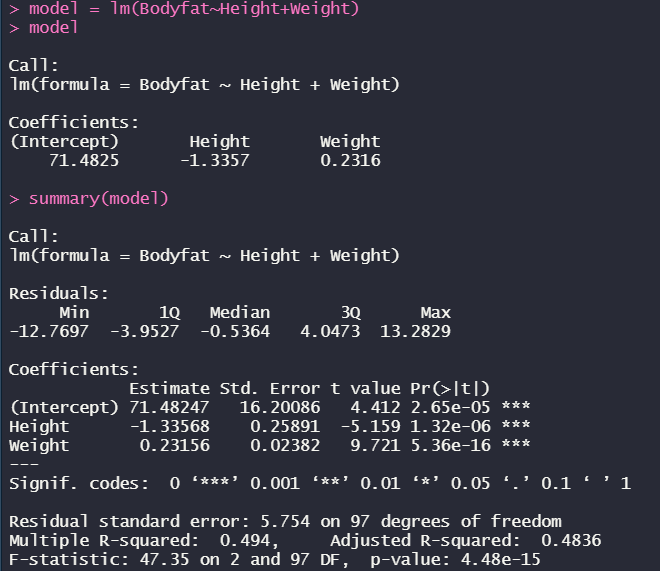
***(X1 = 327.844 – 7.525 \* y)***

1. The percentage of a person’s weight that is made up of body fat is often used as an indicator of health and fitness. However, accurate methods of measuring percent body fat are difficult to implement. One method involves immersing the body in water to estimate its density and then applying a formula to estimate percent body fat. An alternative is to develop a model for percent body fat that is based on body characteristics such as height and weight that are easy to measure. The dataset *BodyFat* in *Lock5withR* contains such measurements for a sample of 100 men.
2. Import the data and Access the variable names included in the dataset.



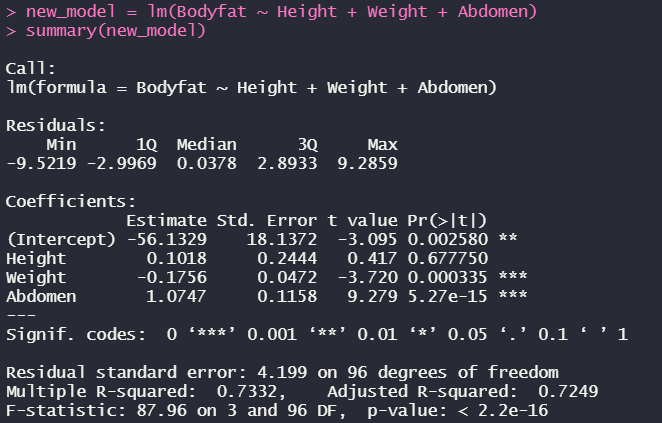


1. Fit a model to predict Bodyfat using Height and Weight. Comment on whether either of the predictors appears to be important in the model.



***(since the p-values for Height and Weight are significantly small, meaning that both them have effect on the dependent variable)***

1. Add Abdomen as a third predictor to the model (b) and repeat the assessment of the effectiveness of each predictor.



***(the p-value of Height gets to be way higher, which means that it is being insignificant with the dependent variable of Bodyfat)***

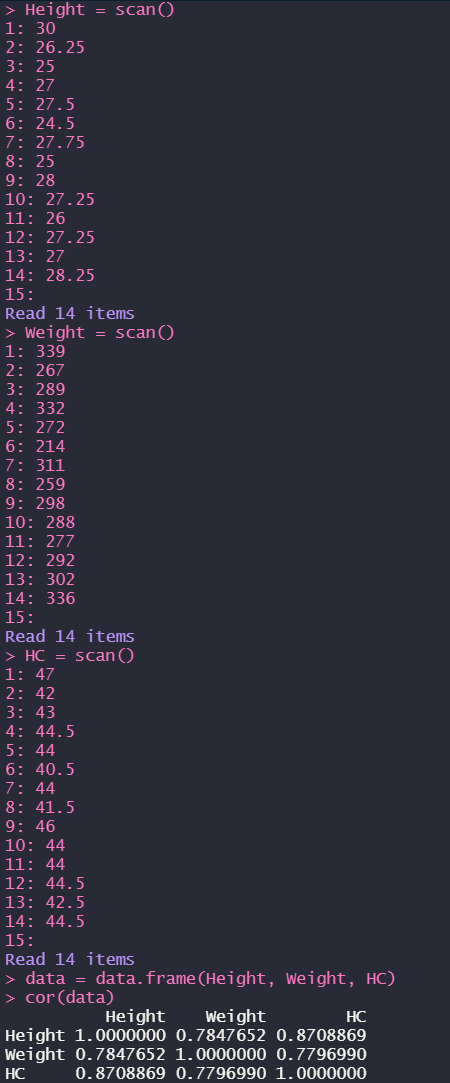
1. Interpret the coefficient of Abdomen you get in the model part (c).

***The p-value of Abdomen is significantly small means that it’s pretty important***

1. A pediatrician wants to determine the relation that may exist between a child‘s head circumference (in centimeters), height (in inches), and weight (in ounces). She randomly selects 14 three-year-old children from her practice and obtains the following data:

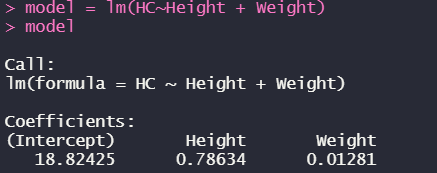
|  |  |  |
| --- | --- | --- |
| Height | Weight | Head Circumference |
| 30 | 339 | 47 |
| 26.25 | 267 | 42 |
| 25 | 289 | 43 |
| 27 | 332 | 44.5 |
| 27.5 | 272 | 44 |
| 24.5 | 214 | 40.5 |
| 27.75 | 311 | 44 |
| 25 | 259 | 41.5 |
| 28 | 298 | 46 |
| 27.25 | 288 | 44 |
| 26 | 277 | 44 |
| 27.25 | 292 | 44.5 |
| 27 | 302 | 42.5 |
| 28.25 | 336 | 44.5 |

1. Construct a correlation matrix. Is there any reason to be concerned with multicollinearity?



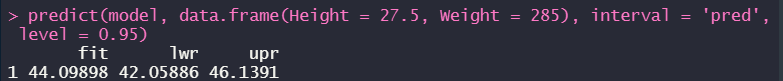
***(Multicollinearity usually happens when independent variables in a regression analysis are correlated, in this example, the p-value are way greater than significance level, meaning that we can support null hypothesis to say that they don’t have correlations)***

1. Find the least-squares regression equation with the response variable, head circumference.



***(HC = 18.82425 + 0.78634 \* Height + 0.01281 \* Weight)***

1. Construct 95% confidence and prediction intervals for the head circumference of a child whose height is 27.5 inches and weight is 285 ounces.



1. Perform the residual analysis of the model.



