Shaan Barkat

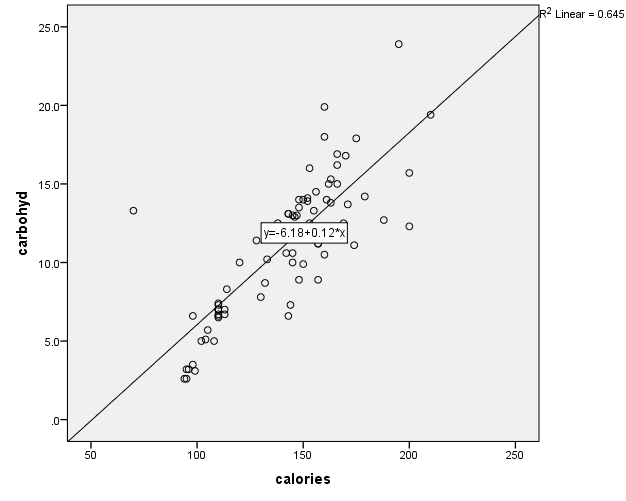
Prof Grady

IT 223

29 January 2017

Assignment #4

1)



1b) Correlation Coefficient = .803

1c) The linear relationship of this scatterplot is strong.

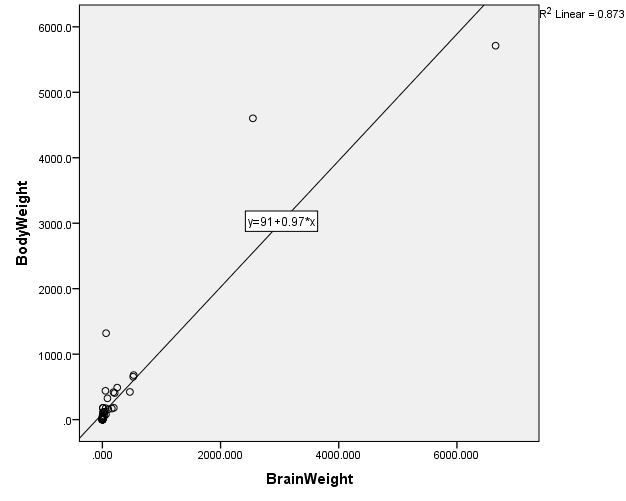
1d) R Squared describes how close the data is to the fitted regression line. In this case, R Squared is .645, meaning the data for the most part is close to the regression line.

1e) y = -6.18 + 0.12\*x

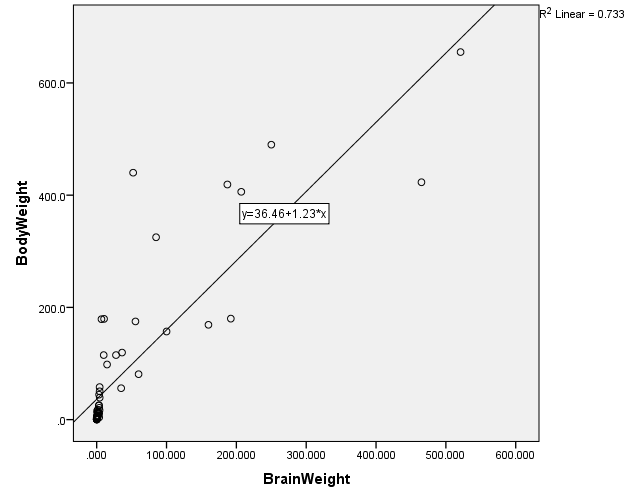
1f) The size of a person can improve the model because this will tell us how much the person weighs, also it tells us the amount of calories that person is consuming per beer. Thus, this will show whether the person is adding weight on because of the beer intake.

2a) There is a huge gap between the majority of the data points and the other two data points. The linear relationship should be strong nonetheless.

y = 91 + 0.97\*x

2b) 

2b) In this scatter plot there are both influential and outliers. This is because the second highest data point on the y axis is considered a outlier because it is in fact the furthest away from the fitted regression line. Also, the data point that is furthest, is in fact an influential data point because it is not too far to impact the regression line but close enough to influence it. (Didn’t know how to word it)

2c) 

Y = 36.46 + 1.23\*x

2d) The biggest body weight was removed and R/R Squared changed because the numbers are smaller. Previously the 4 largest numbers caused an outlier because as the data began to increase in numbers the gap opened wider, but now that the 4 largest numbers are gone, the gap is a lot smaller.