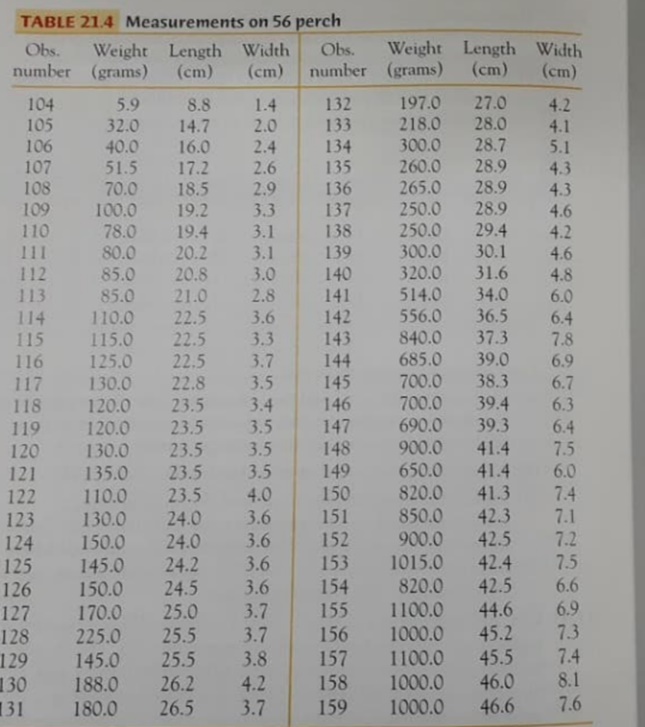
Chad Gian Villanueva - BSCS 3A

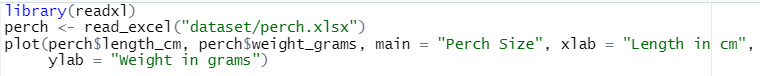
Exploratory Data Analysis Final Project

Table below contains data on the size of perch caught in a lake in Finland. R language will help you analyze these data.

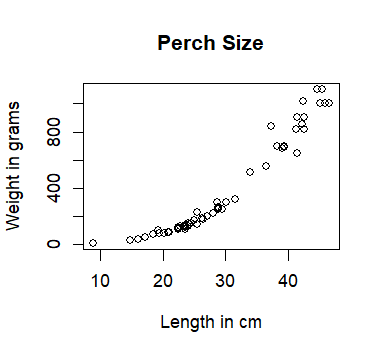


1. Make a scatterplot of weight versus length, with length as the explanatory variable. Could you describe the pattern of the data and any clear outliers?

Code:



Output:



Answer:

The plot is not linear but in a shape of a cone from the lowest number to the highest but curved. As the data rise to the upper right of the plot, it scatters more widely across the plot. Lower left has the most concentrated data but in the upper right the data scatters more. There seems to be only one outlier at the lowest point of the graph.

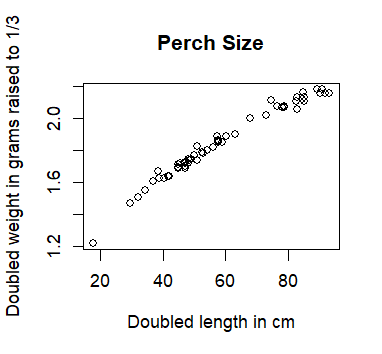
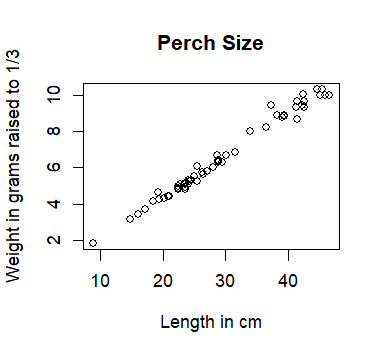
1. It is more reasonable to expect the one-third power of the weight to have a straight-line relationship with length than to expect weight itself to have a straight-line relationship with length. (Hint: What happens to weight if length, width, and height all double?)

Code:





Output:

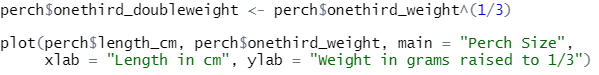


Answer:

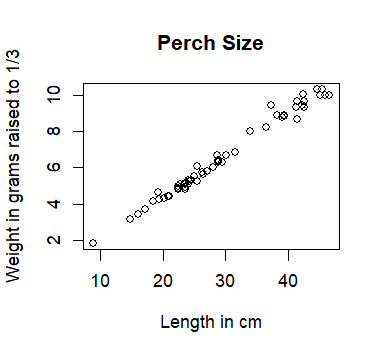
If I double the length and the weight raised to one third, and create a scatterplot out of it, the shape of the scatterplot is curved but to the top unlike in problem a’s scatterplot. By applying the one-third power, the relationship between weight and length becomes more linear. This means that as the length increases or decreases, the corresponding one-third power of the weight changes proportionally in a more straightforward manner, resulting in a straight-line relationship on a scatterplot.

1. Use R language to create a new variable that is the one-third power of weight. Make a scatterplot of this new response variable against length. Describe the pattern and any clear outliers.

Code:



Output:

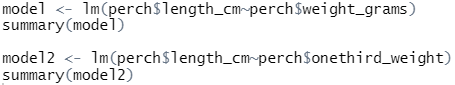


Answer:

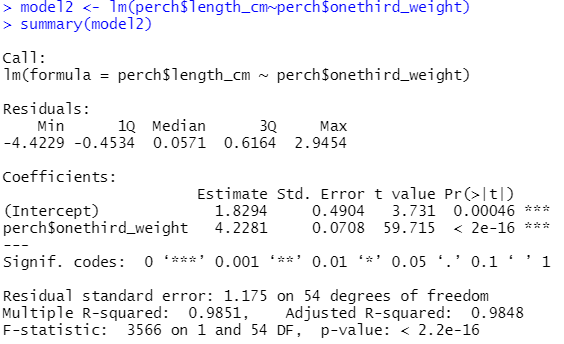
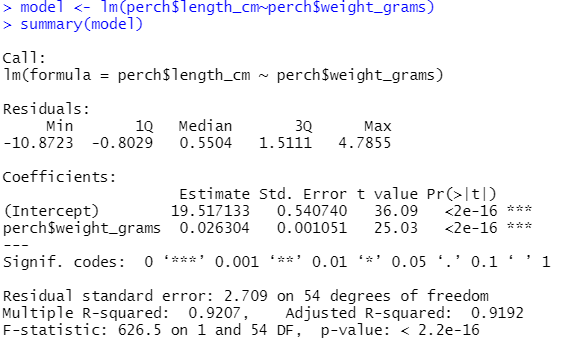
The pattern of the scatterplot is linear has only one outlier at the lower left of the plot.

1. Is the straight-line pattern in (c) stronger or weaker than that in (a)? Compare the plots and also the values of r-squared.

Code:



Output:



Answer:

The straight-line pattern in problem c is stronger than that in problem a. The plot in problem a is a lot more curved than it is in problem c. That means the problem c has more linear scatterplot when compared to problem a. Also, the r-squared for length with weight itself and length with weight raised in one third shows that both models have a strong relationship between the dependent and independent variables. Length with weight raised to one third has a higher r-squared value of 0.9851 compared to length with just weight itself that has a value of 0.9207. This means that when weight is raised to one third, the data can explain a larger proportion of the variance in the dependent variable compared to length and just weight itself. It is simpler to say that when raising the weight to one third, it can provide a better fit to the observed, as it can explain a higher percentage of the variability in the length compared to just weight itself.