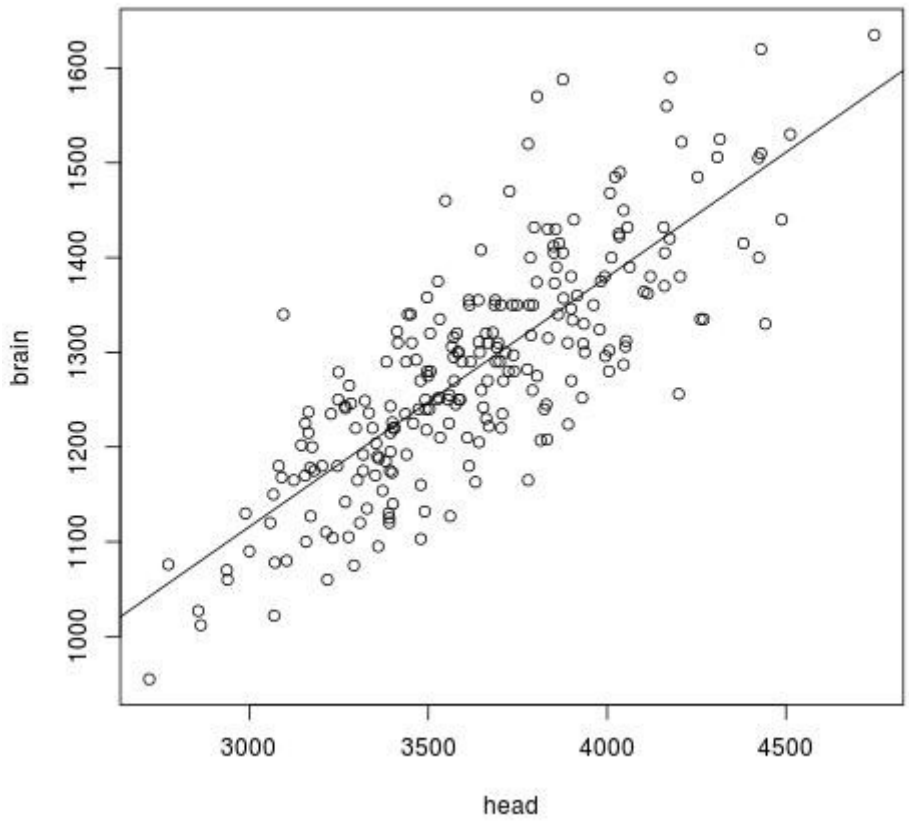
Regression

1PI11CS050-ChitraSingh

1. Generate a scatter plot of the data



2. Fit a linear model to predict brain weight from the head volume using the data

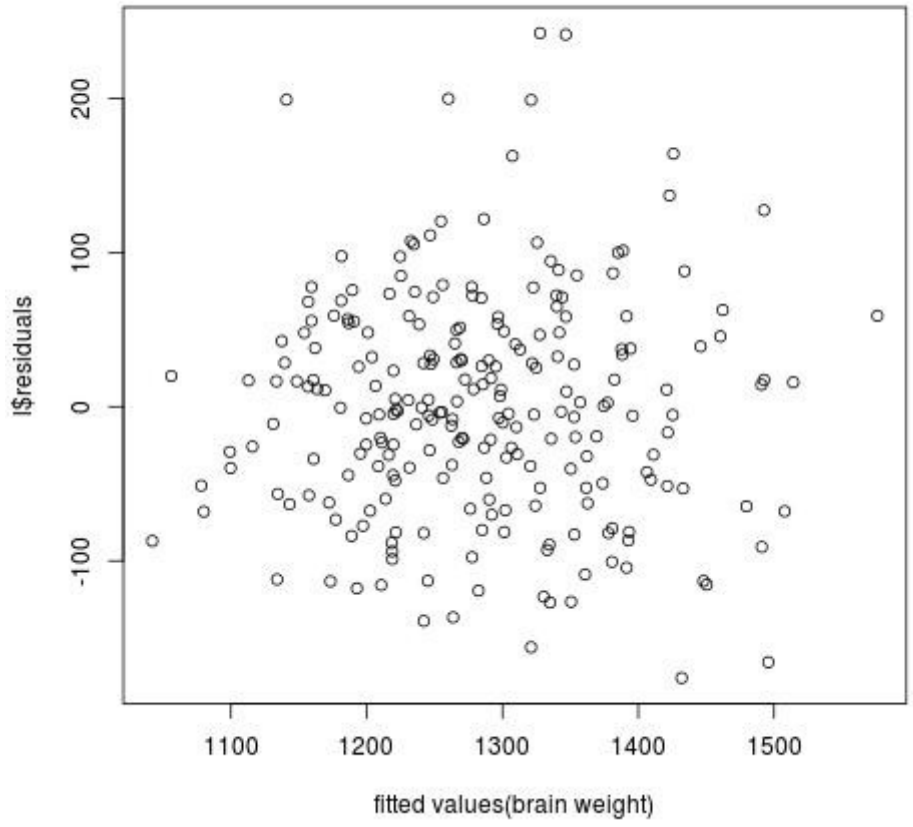
   l$coefficients
   (Intercept) brain
   325.57342 0.26343

   y : brain weight
   x : head volume
   y = 325.573421 + 0.26343 * x

3. Determine the 95% confidence for the values of the regression coefficients

	2.5 %	97.5 %
B_0 : (Intercept)	232.7007553	418.4460868
B_1 : brain	0.2380003	0.2888584

4. Generate a plot of the residuals as a function of the fitted values



5. Determine a 95% confidence interval for the predicted brain weight of an individual whose head volume is 4000 cc

fitted brain weight: 1379.291

lower interval: 427.9517

upper interval: 2330.63

Results and observations:

1. Based on the scatter plot, is a linear model suitable?

Yes, this is linear model as the points in the scatter plot are evenly distributed around the line predicted. No non linear association ca be seen.

2. Based on the residual plot, is a linear model suitable?

The scatterplot is linear can further be justified by the residual plot of residuals v/s fitted values. There is no **substantial pattern to the plot**, and the vertical spread doesn't vary too much (except two points on the either side), hence it can be concluded that this is a linear model.

This can be stated because from a mathematical point of view, residuals have a mean of 0, and the correlation between residuals and fitted values is 0 as well. Least squares line is therefore horizontal.