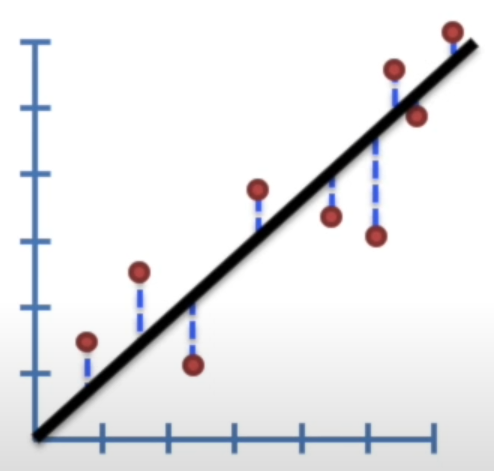
Linear Regression

Main ideas of linear regression:

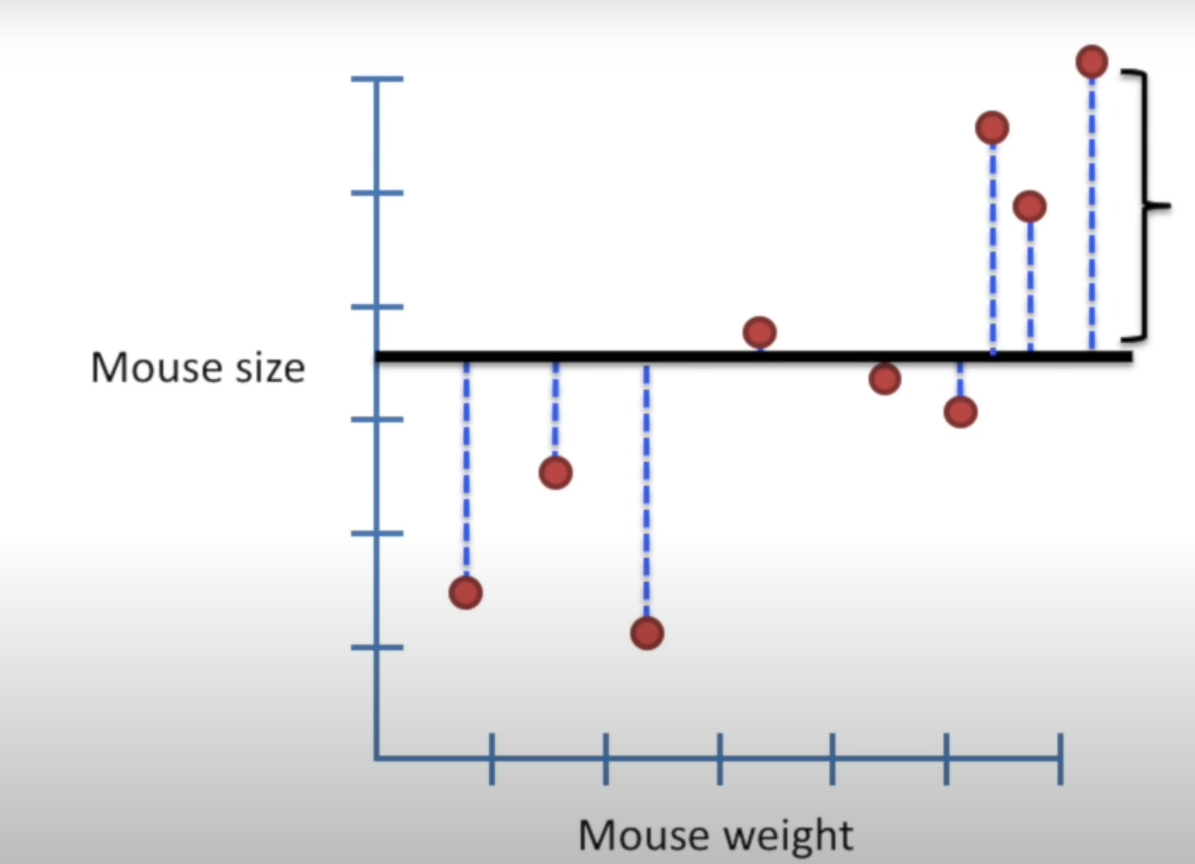
1. Use least square to fit a line to data.



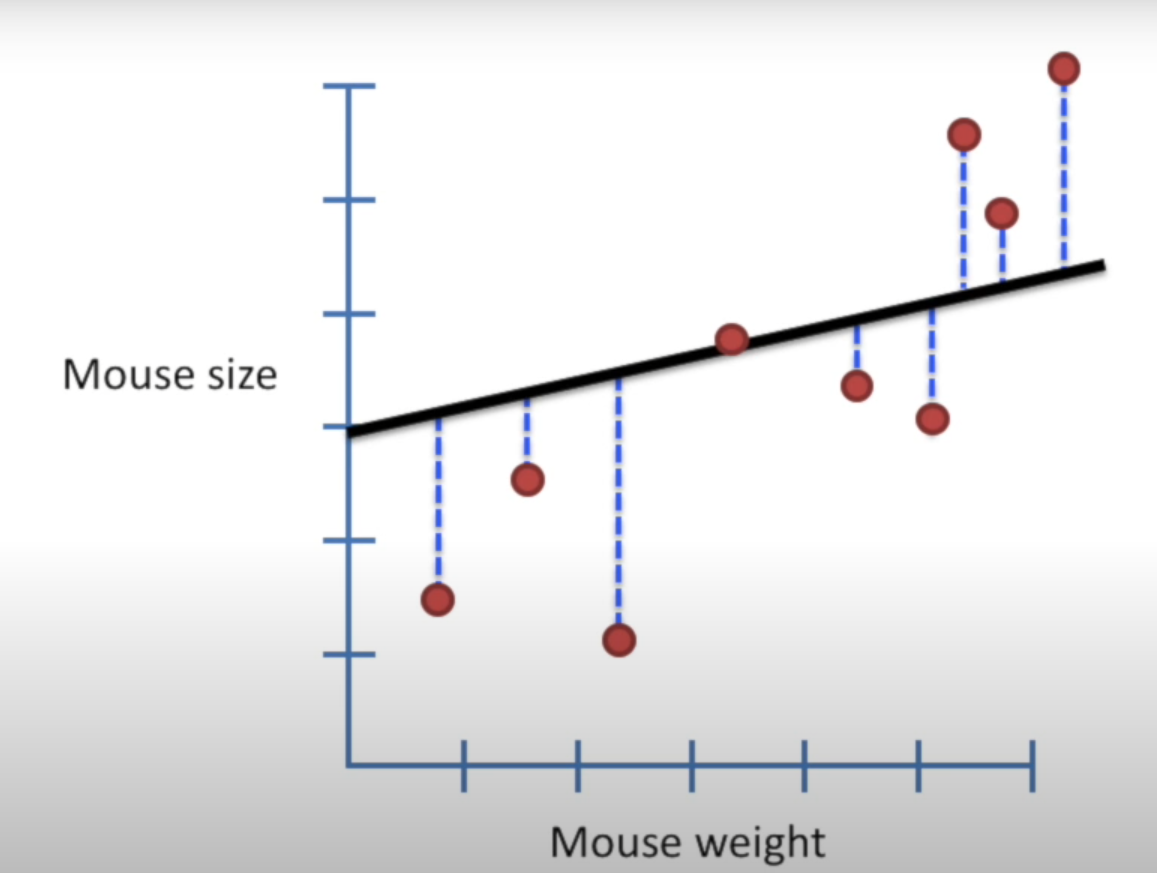
1. Calculate R2
2. Calculate p-value for R2

Steps:

1. Draw a line through the data.
2. Second measure the distance from the data to the line, square each distance and add them up. The distance from the data point to the line is called the residual.



1. Now rotate the line and then again find the residuals and square each residual and sum them up.

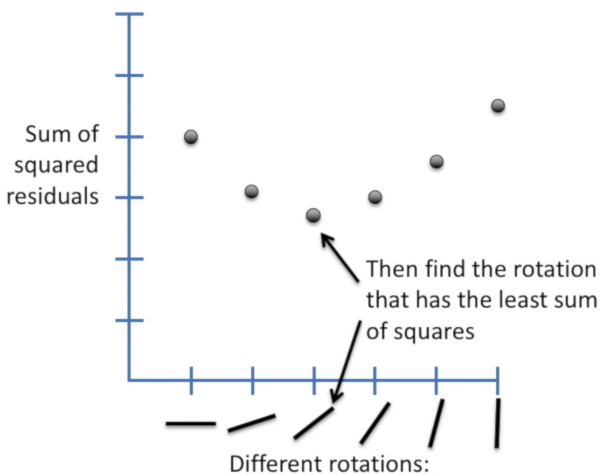


1. Now again rotate the line and square the residual and add them up for different slope and intercept values.

Y = mx + c

Here m is the slope and then c is the intercept. This is a linear equation of an straight line.

1. Find the line which give less sum of squares value (residual value)



1. Calculate the R-Square value

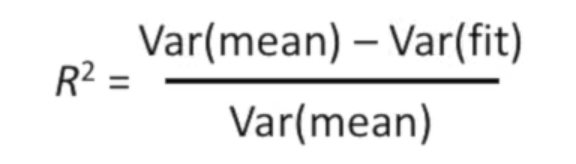
The R² (coefficient of determination) tells us how well the fitted line **explains the variance** in the dependent variable relative to the mean model.

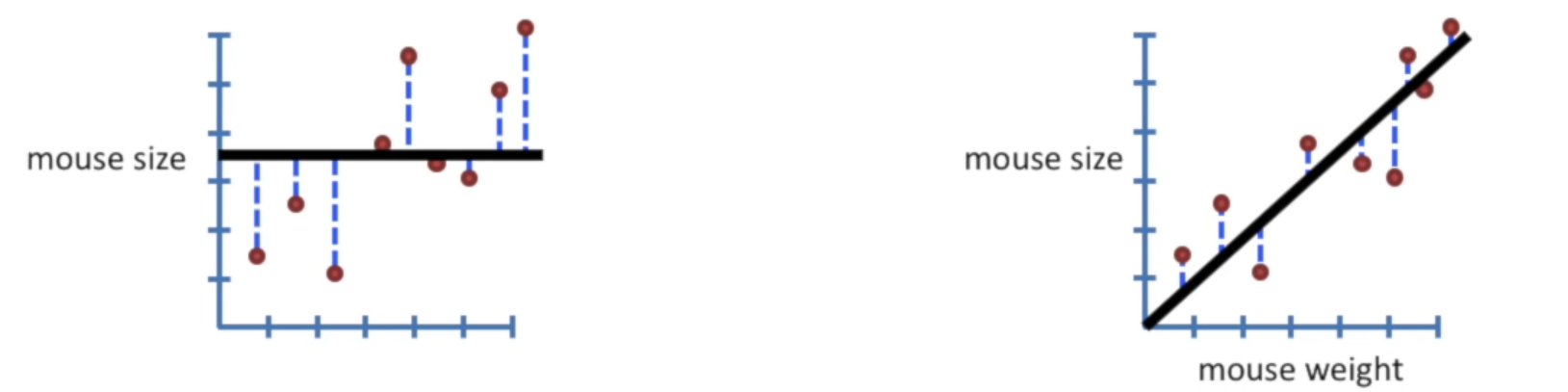
SS (mean) = (data – mean)2

Var (mean) = SS (mean)/n

SS (least squares fit) = (data – line)2

Var (Least square fit) = SS (least squares fit) / n





* The more the parameters you add, the variance around the fit decreases and R2 value get increase, which is not good. Even though the new variable is not that correlated with the dependent variable it still gives a better R square value. Therefore, we have to check the significance of R2. For this we have calculate a p value or (F value)