Linear Function Modeling :-

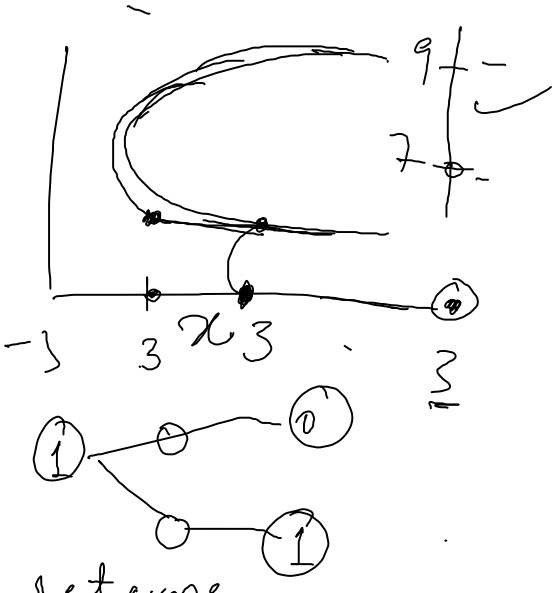
Function

Input → (fn) → Output

3 — 9
3 — 7.

Important note: Fn must return ~~one~~ unique output for the given input.

x → [~~factorial~~] → x!



Function can be defined using :-

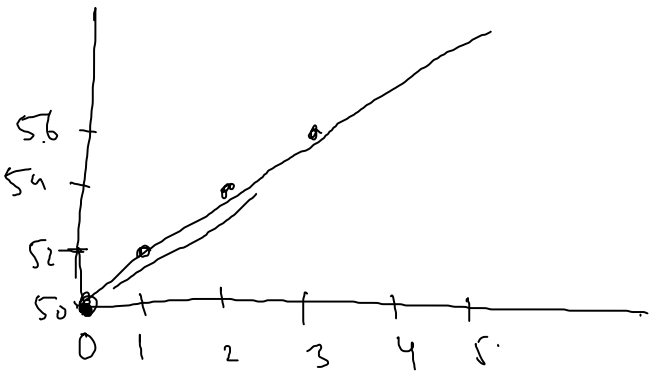
- Words
- Table
- Graph
- Formula / Expression.

Words :- For integer input, a system (program) returns double of the input value, starting from 50

Table :-

Input	Output
0	50
1	52
2	54
3	56

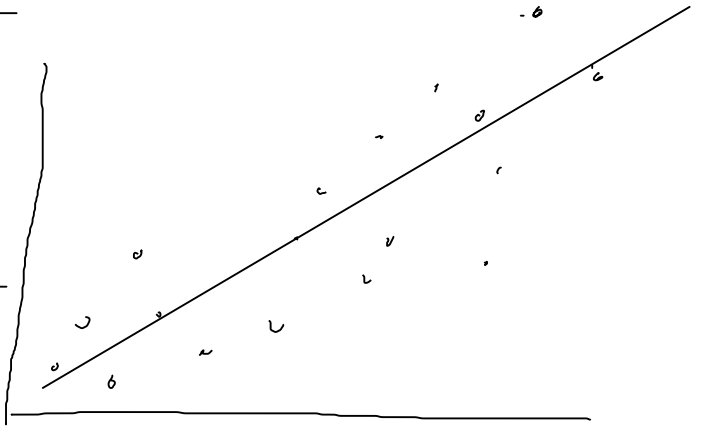
Graph



Formula :- $y = 50 + 2x$

In this case function is a linear fn.

Actual Data :-



Seems like linear relationship.

Can we model the relationship?

We can use R to find the "r" (linear correlation), which would be quite strong in this case (0.91).

We can do more by giving our reader more info than that, for that we need model.

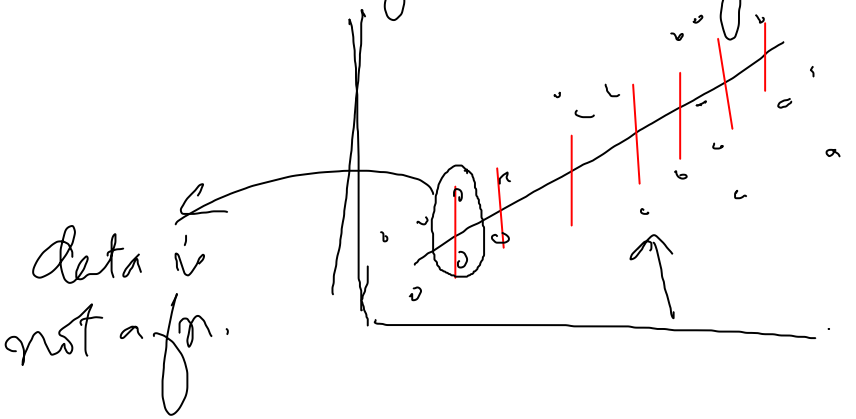
Let's find linear model for this data.

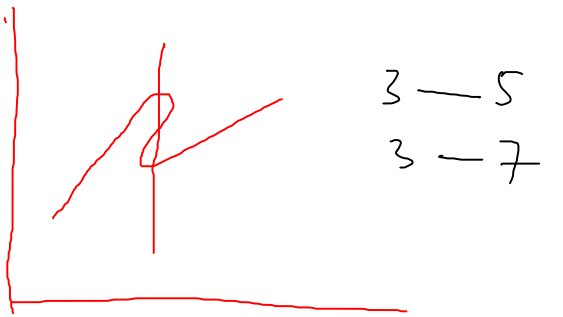
Let's say we come up with the formula

$y = 50 + 2x$, then it says that to predict y start at a line at (0, 50) and then for every int increase

in x increase 2 times of x.
So as per definition of the function, we can have only one output for the given input.

This is normally tested by "vertical line" test.

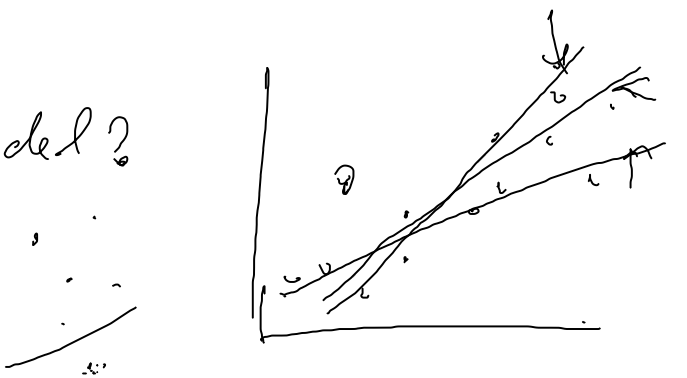




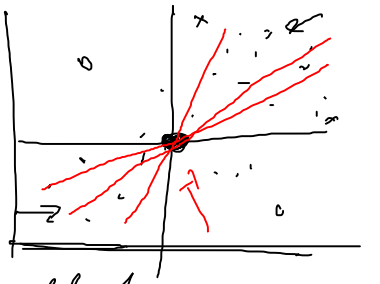
Our aim is to find the relationship in data, so that our model can explain that what is the relationship look like.

Line of Best Fit :-

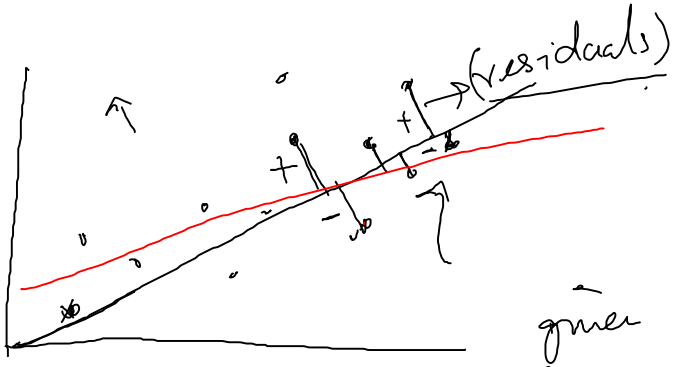
How can we decide the model?
There are ~~two~~ ∞ possibilities.
Q:- what is the best fit of the data?



~~Any~~ Best fit must pass through the mean_x and mean_y.



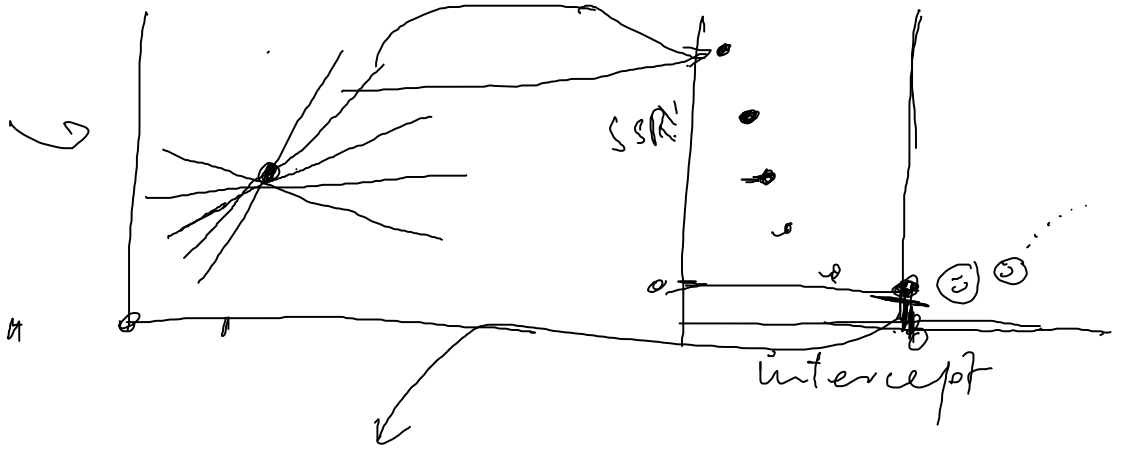
Still there are ∞ possibilities, therefore we need a new technique, called it is called "residuals"



Formally residuals are defined as $(y - \hat{y})$ — y: given, $\hat{y}$: predicted

Some points lie on top of the line while others lie below the line. Collective effect of the residuals will be cancelled out. For this reason we shall take sum the squares of the residuals, (SSRV).

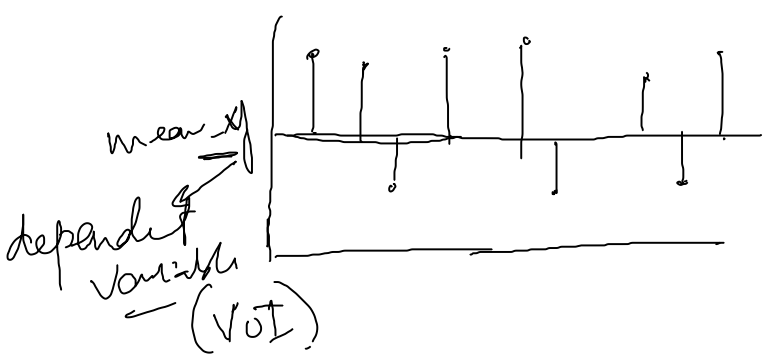
This value helps us to find out the appropriateness of the given line. The model with the smallest SSRV will be considered as the best fit.



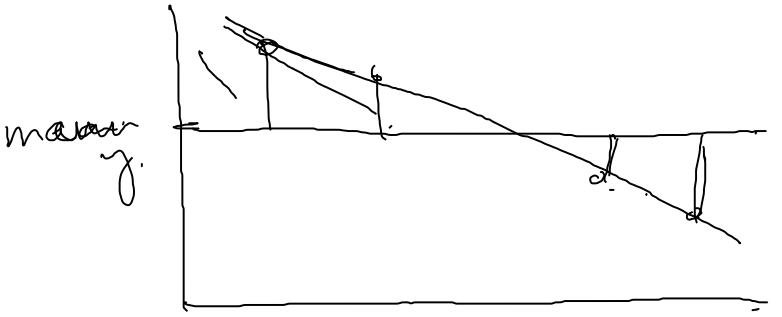
$y = \text{intercept} + \underline{\text{Slope}}\, x.$

→ Sum of squared residuals is called "coefficient of determinat". It is ~~Cof~~ closely related to "r", and it computed by squaring r. ⇒ $\underline{r^2}$ → (Pearson coefficient)

$r = \underline{0.9}$, $r^2 = \underline{.81}$



(Sum of squared residuals) → Denominator



Sum of Squared predicted values − the mean y,

by taking the ratio

numerator

num / deno : R^2

0.99