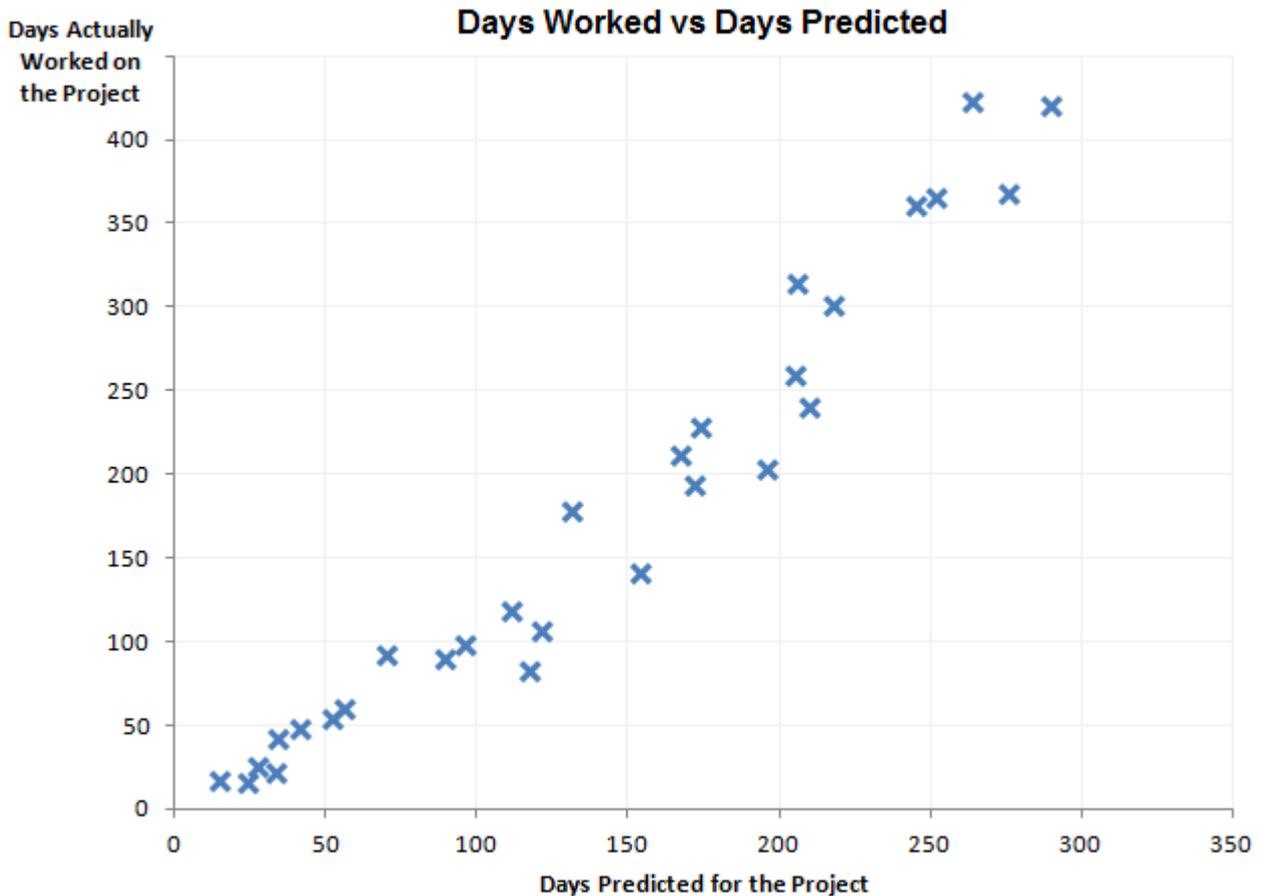


## Level 1 Data Practice #11

Alphaville Computing is interested in how well they predict the number of work days it will take to complete a project.

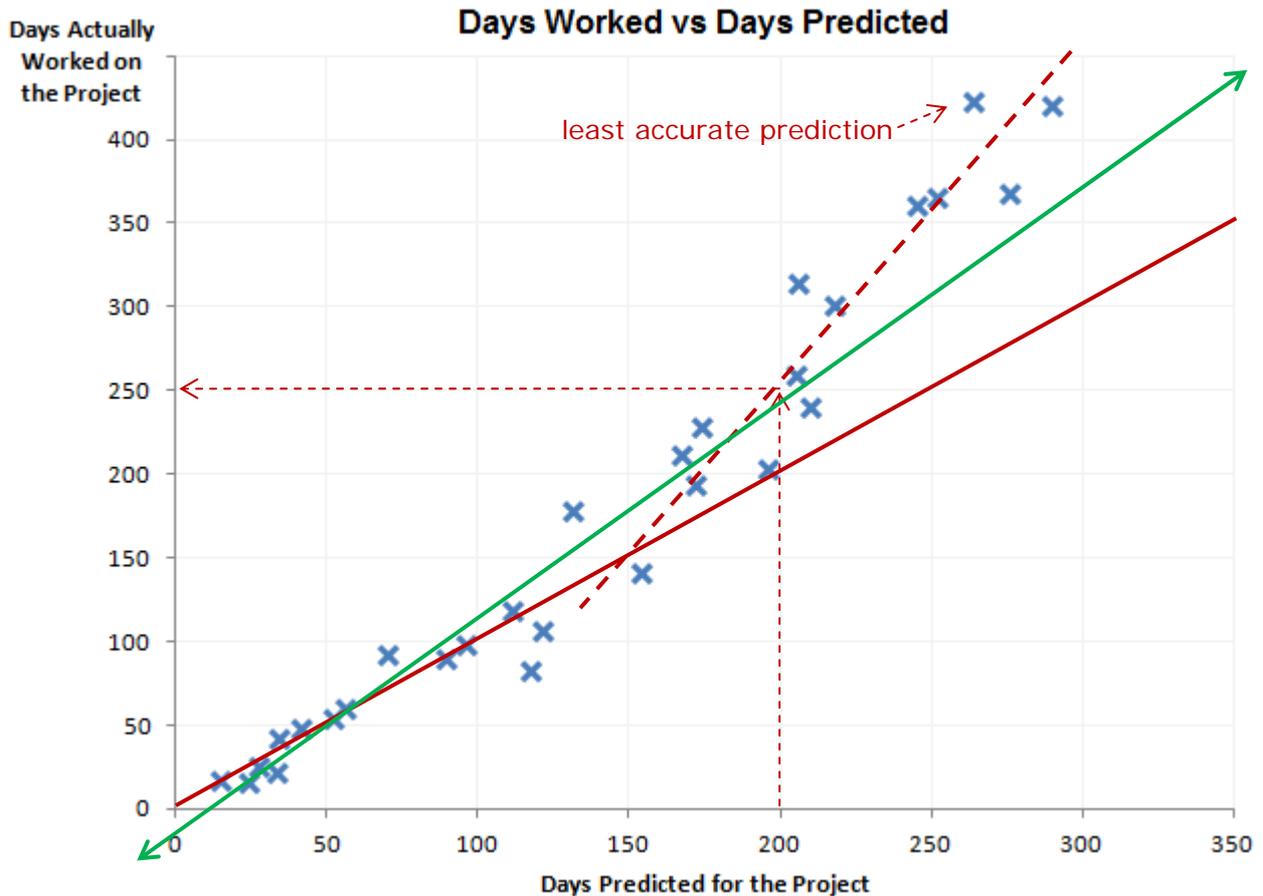
If their predictions are too low they will tend to bid too much, and will lose business to cheaper bidders. If they predict too little they will win lots of projects, but will lose money, as they will work for days without being paid.

The results of all the projects started in a year are shown below.



1. How many projects did they start during the year?
2. Describe, fully, how accurate they were at predicting how many days a project would take.  
Show on the graph how we can visualise their accuracy in predicting.
3. How would you describe the relationship between the days predicted and the days taken?
4. If during that year they had predicted that a project would take 200 days, how long do you think it would probably have taken?  
How reliable do you believe your answer is?
5. Construct at least two different models for the data shown, and say which you think is better and why.

## Answers: Level 1 Data Practice #11



- They started 29 projects (the number of dots shown).
- If the predictions were accurate, the dots would all be on the line shown in solid red.  
 Instead we see that they were quite accurate up to 150 days but after that point their accuracy dropped off steadily and projects took a lot longer than expected. The worst example is 150 days above the solid red line.  
 For longer projects the dotted line shows after 150 days three days are worked on average for every two predicted.
- There is a strong positive linear correlation for the first 150 days (solid line), then there is a less strong positive linear correlation with a different relationship (dotted line).  
 (If instead a student is using the green line of overall best fit: there is an overall moderate positive correlation, but that the data is probably not linear overall as there is a pattern in the spread around the line of best fit.)
- They would likely take 250 days to finish a project predicted to take 200 days (using either dotted red or green makes little difference here. The value is probably accurate to about  $\pm 50$  days, as that is the scatter we see at that point either side of the lines).
- 1)  $\text{Actual} = \frac{350}{450} \text{Predicted} - 10$  (or close equivalents) using one overall best fit (green)
  - 2)  $A = P$  for  $P$  up to 150; then  $A = \frac{300}{200} P - 75$  for  $P$  over 150 when using two line fit (red)
 The two line fit is much better, as the scatter from it is much less and without a pattern.  
 (A curved fit is technically possible but it's too hard to fit a formula with only Y11 Maths.)