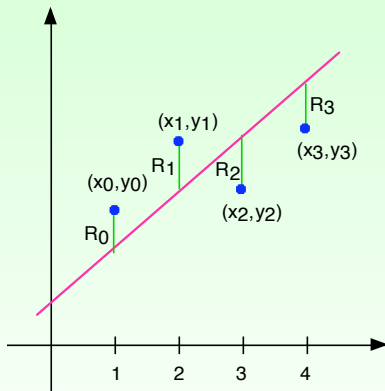


## A standard statistics problem:

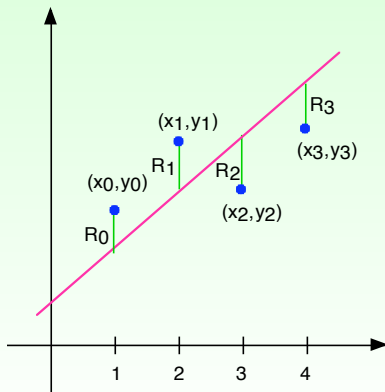
Given a bunch of data points, find a line that minimizes the total (vertical) distance between those data points and the line.



(For each  $x_i$ , look at the distance between the  $y$ -value predicted by the line, and the actual  $y$ -value of the data.)

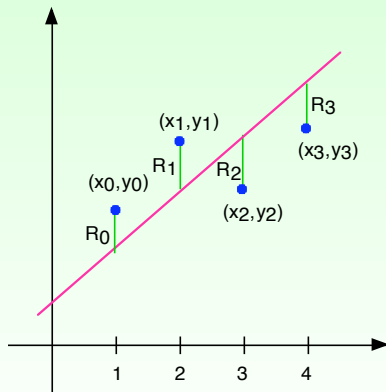
## A standard statistics problem:

We want to minimize the overall distance of the points from the line, using these vertical distances, called **residuals**. Since we don't want negative distance to cancel out positive distance, we will instead minimize the sum of the squares of these *residuals*.



## A standard statistics problem:

The line that minimizes the sum of the squares of these *residuals* is called the **linear regression** line.



Archaeopteryx is an extinct beast that is generally regarded to have been the first bird. Here are the lengths in centimeters of the femur and humerus for five fossil specimens that preserved both bones.

Humerus ( $x$ )	41	63	70	72	84
Femur ( $y$ )	38	56	59	64	74

1. Set up the sum of the least squares that we need to minimize to find the least-squares regression line.
2. Find the regression line. (Feel free to use Maple if you have your a laptop with you; you can mimic what I did on the first Maple file of the day if you want.)
3. If an archaeopteryx had a humerus of length 69 cm, how long would you predict that its femur was?