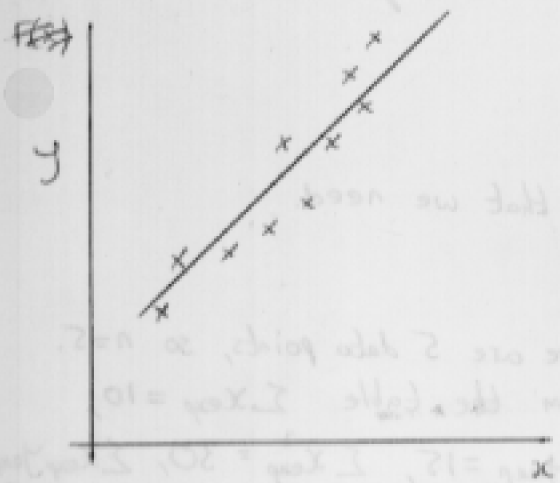
Tutorial 7



If we did an experiment, we would ~~obtain~~ take a series of measurements and end up with a series of data points.

When we plotted the points, in all likelyhood, they wouldn't be in a straight line. However, we might be interested in finding an equation for a straight line that best approximates the data.

The equation of a straight line is $y = ax + b$, so it becomes a question of finding the most suitable values of a and b.

If the points were all in a straight line, then we could pick a & b so that, at each value of x, the experimental value of y minus the equation value of y is zero. Even if they weren't in a straight line, we could try and find a and b so that they minimize the value of the subtraction. This might give us the best fitting line.

To do it for all the points, we would have to include a summation sign. Effectively, we would try to minimize $\sum_{\text{data points}} (y_{exp} - y(x_{exp}))$ (where $y(x_{exp}) = ax_{exp} + b$), by getting the right values of a & b.

However, this approach has a problem. Imagine a line with ~~the~~ a negative gradient that passes through one of the points on the graph. Obviously, a line with a negative gradient would be a bad fit for the data points. But, if, for every point above the line, there was another point, an equal distance below the line, $\sum (y_{exp} - y(x_{exp}))$ would still be zero. Instead, we use $\sum (y_{exp} - y(x_{exp}))^2$, which, because of the squaring, doesn't let points the same distance above and below the line cancel each other out.

The values of a and b that minimize $\sum (y_{exp} - y(x_{exp}))^2$ are

$$a = \frac{n(\sum x_{exp} y_{exp}) - (\sum x_{exp})(\sum y_{exp})}{n(\sum x_{exp}^2) - (\sum x_{exp})^2}$$

$$b = \frac{1}{n}\left((\sum y_{exp}) - a(\sum x_{exp})\right)$$

where n is the number of data points, $\sum x_{exp}$ a sum over the x values, $\sum y_{exp}$ a sum over the y values, $\sum x_{exp}^2$ a sum over the square of the x values and $\sum x_{exp} y_{exp}$ a sum over the product of the