

4.5.1: Graphic teaching aid for regression and correlation

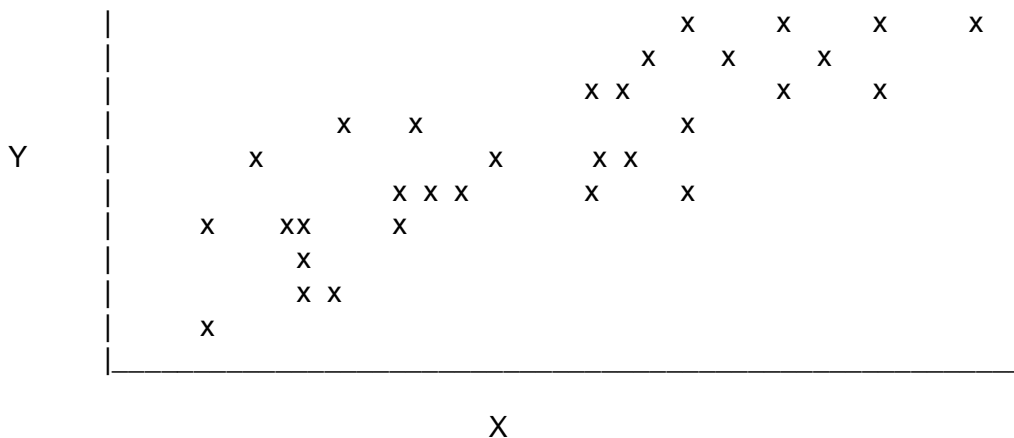
[First published, 7 March 2015: revised **draft only** 28 November 2018]

I've been using SPSS since 1972, but was totally new to this sort of thing in 2005 when updating teaching materials. I tried to reproduce my classroom demo on regression and correlation. My students tended to be from sociology and related areas, and generally had little or no previous experience of computing or statistics.

SPSS was never intended to do this next bit: it probably needs a professional animator. Open University in UK has some good animations for stats. I'm a lot more fluent now in Windows and Word and also have a scanner to produce jpg or pdf files, but here's what I attempted at the time.

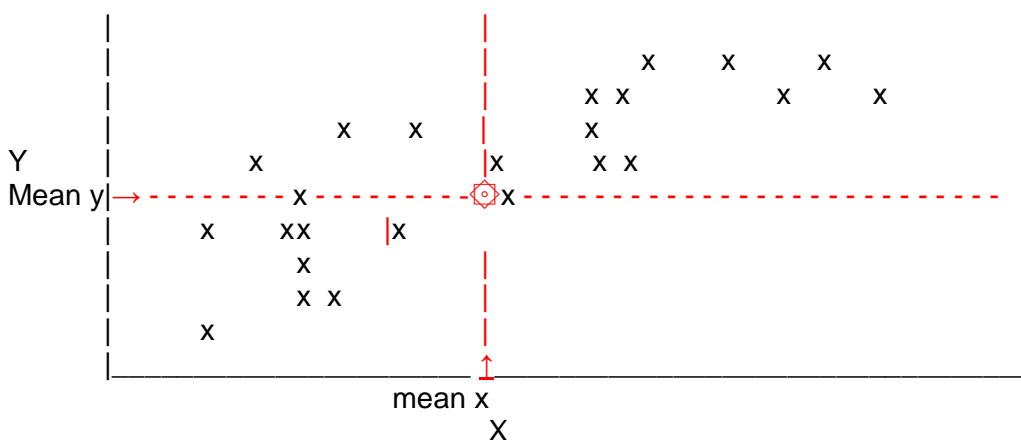
Anyway I started off in Word:

Fig 1: Original handmade scatterplot



Draw vertical line through mean x and horizontal line through mean y

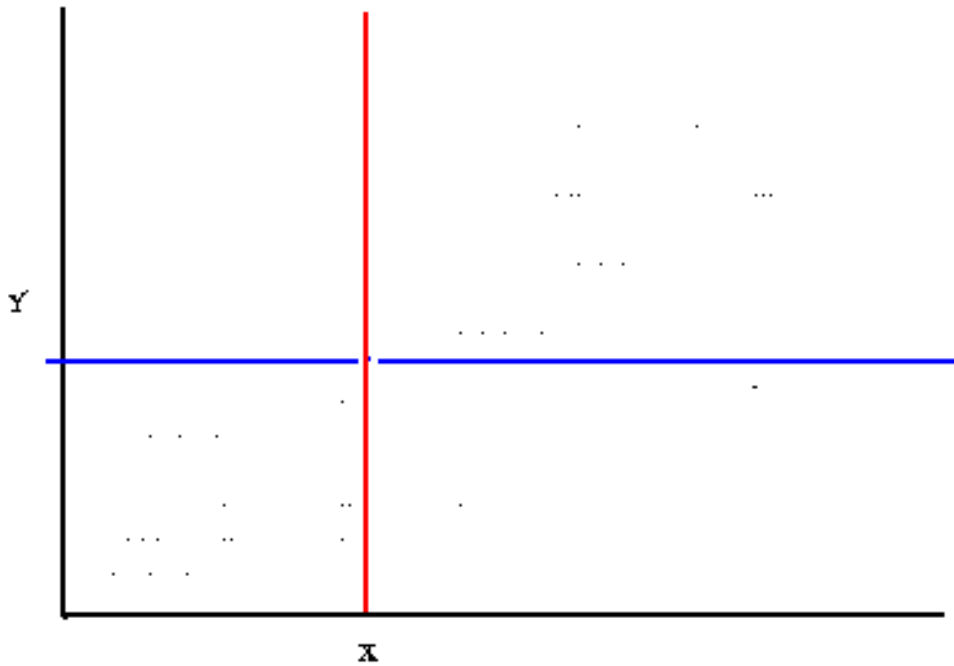
Fig 2:



⊙ marks the intersection of mean x and mean y

Fig 3:

Vertical red line = mean of X  
Horizontal blue line = mean of Y




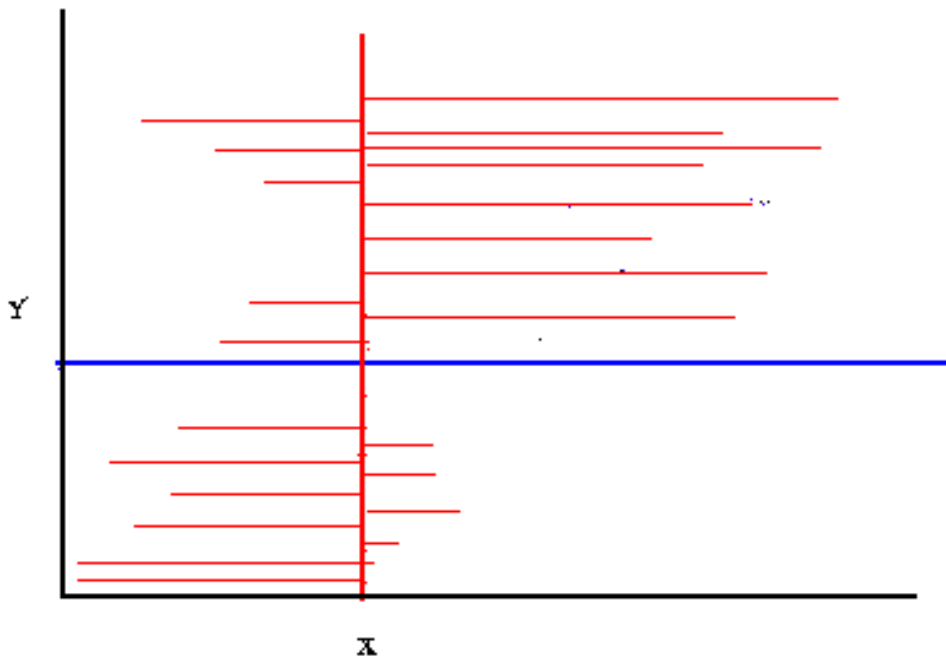
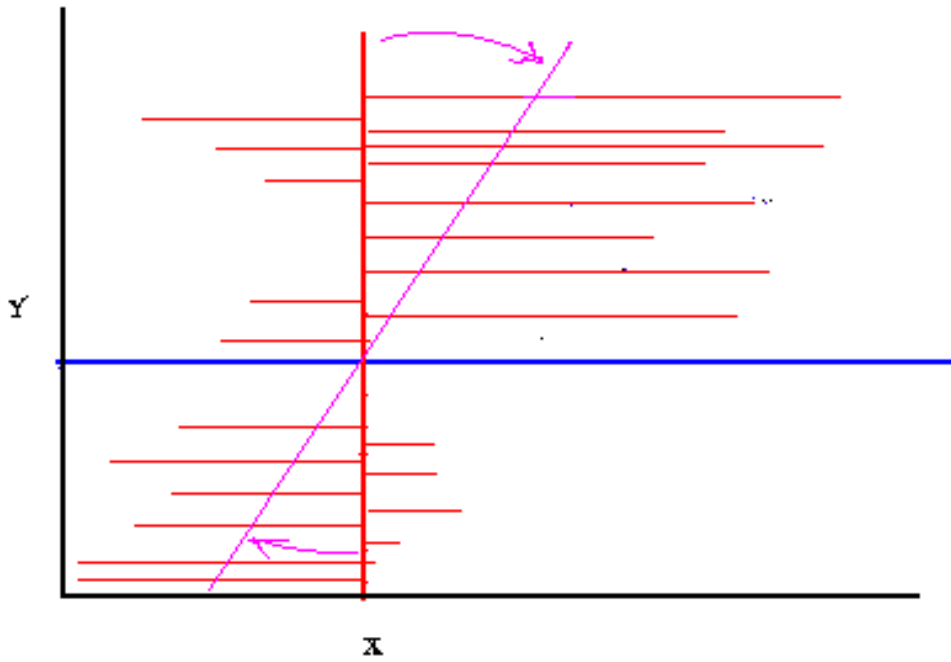
Imagine a) the scattergram is a flat wooden board with pegs marking each data point and b) that the vertical red line is a rigid pole with a hole in it and a nail hammered through it into the board at the intersection of the means  so that it is free to rotate. Hold the pole steady at vertical and attach elastic bands horizontally from each data point to the pole.

Fig 4:



What will happen if I let go of the pole? Yup, even social work students get this one! They even guess where the pole will stabilise.

Fig 7: Regression of X on Y (violet line)



After that it's easy to explain that the **violet line** actually represents the **regression line of x on y**.


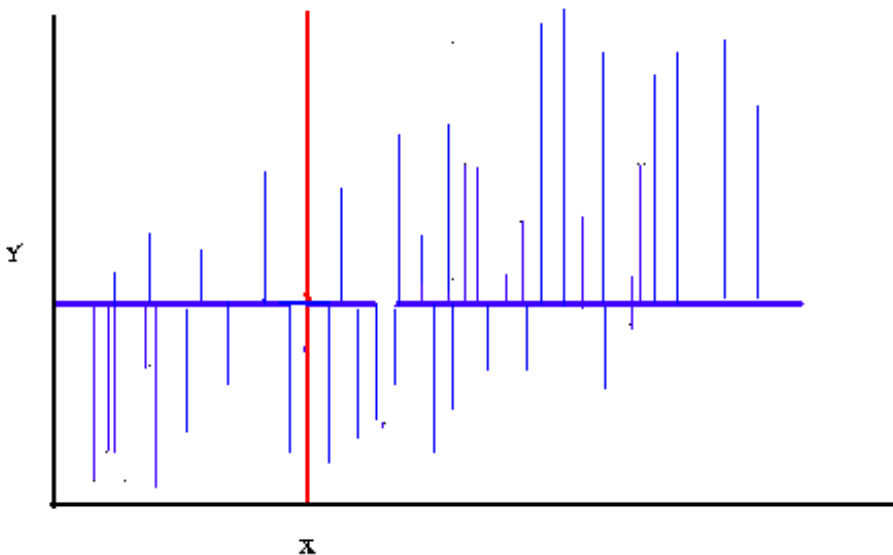
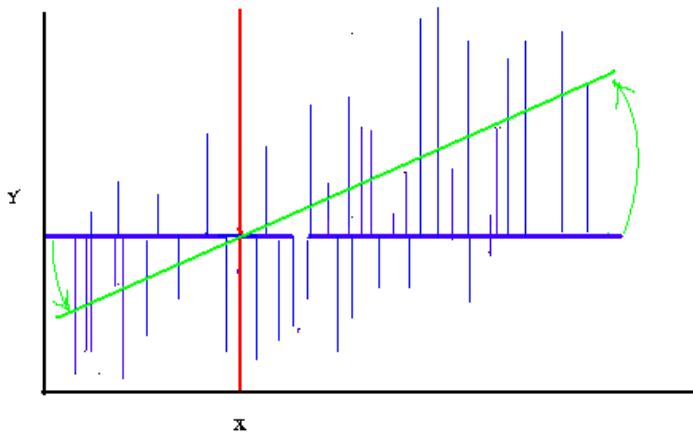
Imagine a) the scattergram is a flat wooden board with pegs marking each data point and b) that the **horizontal blue line** is a rigid pole with a hole in it and a nail hammered through it into the board at the intersection of the means  so that it is free to rotate. Hold the pole steady at horizontal and attach elastic bands **vertically** from each data point to the pole.

Fig 8:



What happens when I let go of the pole? Yup, right again!

Fig 9: Regression of Y on X (green line)



Students easily understood that the pole stops rotating when all the tensions balance out (green line) They can also guess where it will stop.

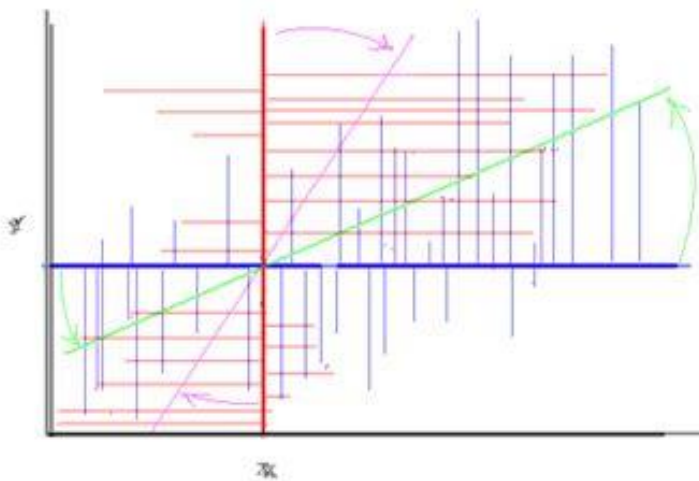
After that it's easy to explain that the green line actually represents the **regression line of y on x**.

Somewhere in here I used the diagrams to explain why you have to square the distances from each data point to the mean before adding them together, otherwise the negative ones cancel out the positive ones, before calculating variance and standard deviations. Going step by step through the process, one data point at a time, helped to build up an equation (very gently).

I'm not sure if the mathematics of elastic is exactly the same for minimising sums of squares.

Pearson correlation  $r$  is the **cosine** of the angle between the green and purple lines. If you overlay the two diagrams with the green and purple lines, you can explain correlation in terms of the cosine of the angle between the two regression lines. Cosine  $90^\circ$  is 0, cosine  $0^\circ$  is 1, *et voilà!*

Fig 10: Overlay showing both regression lines



I never used artificial data in any of my classes. All examples were drawn from professional quality surveys (even small data sets sampled cases from full scale surveys) or from official published sources (eg health, crime and similar statistics).