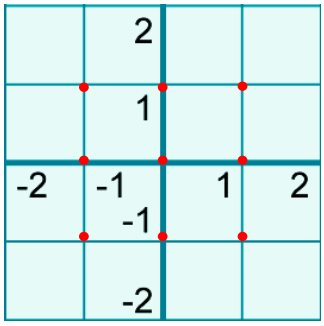
Solutions to Statistics 8 Tutorial: Correlation & Linear Regression

Activity Can you draw the least square line for a set of points without computing the line? Consider the set of 9 points (0,0), (0,1), (0,–1), (1,0), (–1,0), (1,1), (1,–1), (–1,1), (–1,–1) shown in the diagram below.



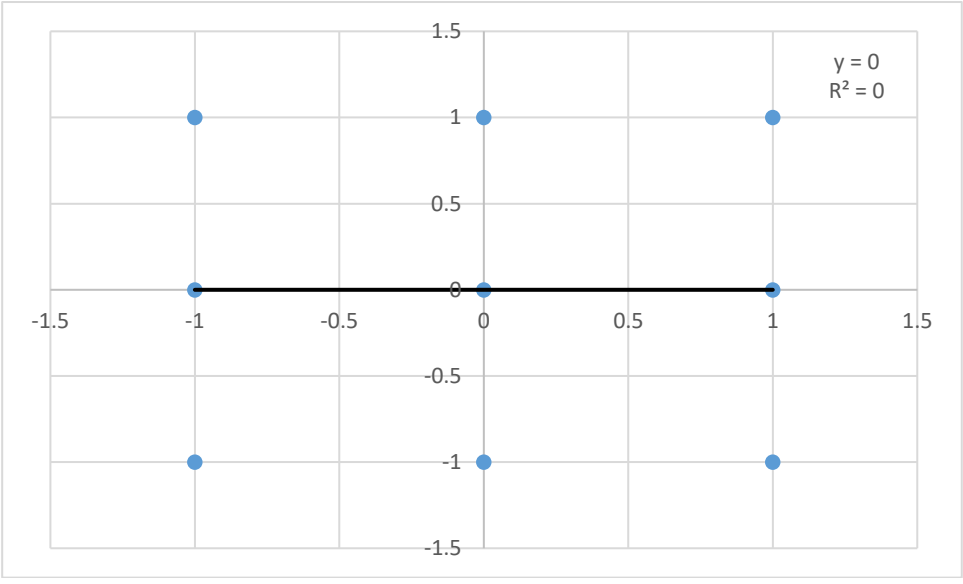
What do you think is the least square line for these 9 points? Draw it on the diagram.

Now, try obtaining the least square line using the the GeoGebra applet:
https://www.geogebra.org/m/t5nbhyBf

Questions to ponder:
1. What do you observe between the line you have proposed and the one obtained using the Geogebra app?
2. Why do this happen?

Further questions to ponder:
1. For any set of points, is it always possible to have a least square of zero?
2. When will this happen?

Solution

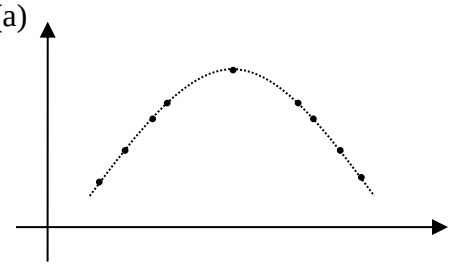


Basic Mastery Questions

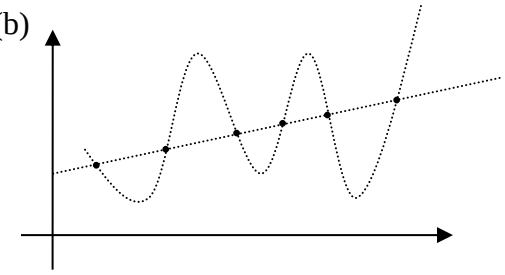
1. Sketch scatter diagrams indicating the following:
   (a) A linear (product-moment) correlation coefficient close to zero but an obvious relation between the variables.
   (b) A non-linear relation between the variables yielding a correlation coefficient of +1.

Solution:

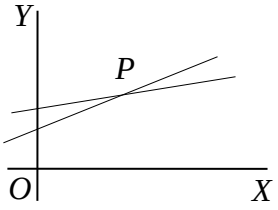
1. (a)



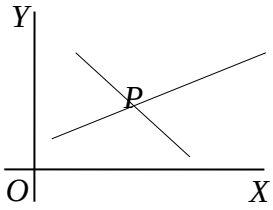
(b)



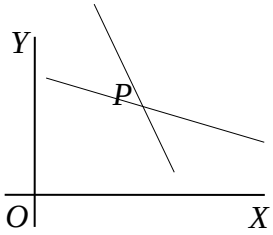
2. The diagrams below show the two regression lines (y on x and x on y) for three different bivariate distributions. The scales along the two axes OX and OY are the same in each diagram.
   Explain, in each case, what the diagram tells us about the correlation between the variables X and Y. What does the point P represent?



(i)



(ii)



(iii)

Solution:

2. (i) X and Y have a strong positive linear relationship.

   (ii) Impossible diagram as the gradient of the 2 lines must be of the same sign.

   (iii) X and Y have a weak negative linear relationship.

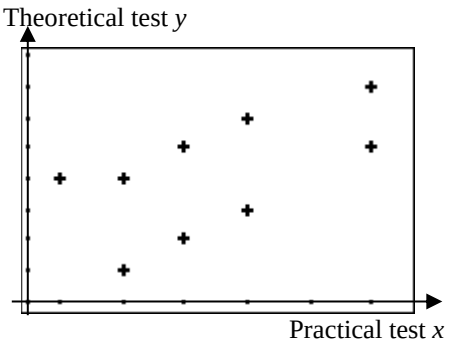
   The point P represents $(\overline{x}, \overline{y})$.

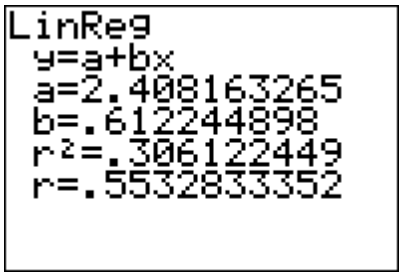
3. Ten students sat for two physics tests, one practical and the other theoretical. Their marks out of 10 are recorded in the following table.

Practical test	8	6	10	8	5	6	8	10	7	7
Theoretical test	6	7	8	6	7	4	9	10	5	8

Draw a scatter diagram of the pairs of marks.
Calculate the product moment correlation coefficient for the data.
A student was absent from the theoretical test but obtained a mark of 6 in the practical test.
Use the appropriate regression line to estimate a mark in the theoretical test for this student.
Comment on the reliability of this estimate.

Solution:





From GC, $r = 0.553$.

Let x = mark of practical test, y = mark of theoretical test

Use regression line of y on x: $y = 2.4081 + 0.6122x$.
When $x = 6$, $y = 6.08$.
This estimate is NOT reliable as the value of r (not close to ± 1) and the scatter plot does not indicate a strong correlation between the practical and theoretical test marks.

4. 2010/NYJC/II/7

An experiment was conducted to investigate the relationship between the amount of unreacted chemical, x, and the time that elapsed since the start of the experiment, t.

x	25.5	28	31	33.5	41	43.5	45.5	51	57.5	58	73
t	46	44	35.5	30	20	15.6	17	12.3	11	8.3	6

If there was an error in recording the t values and all the t-values must be increased by 3, what would be the effect on

(a) $\bar{t}$,

(b) standard deviation of t,

(c) the correlation coefficient, r?

Solution:

(a) $\bar{t}$ will increase by 3

(b) standard deviation of t remains the same

(c) r remains unchanged