Question

(20 points)

You have collected the following data:

x	y	xy
550	53	
590	47	
640	43	
660	40	
360	58	
350	57	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

Answerlist

- Complete the table.
- Calculate the correlation coefficient (r) using the formula below.

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

- The least-squares regression line will be represented as $y = a + bx$. Determine b and a using the formulas below.

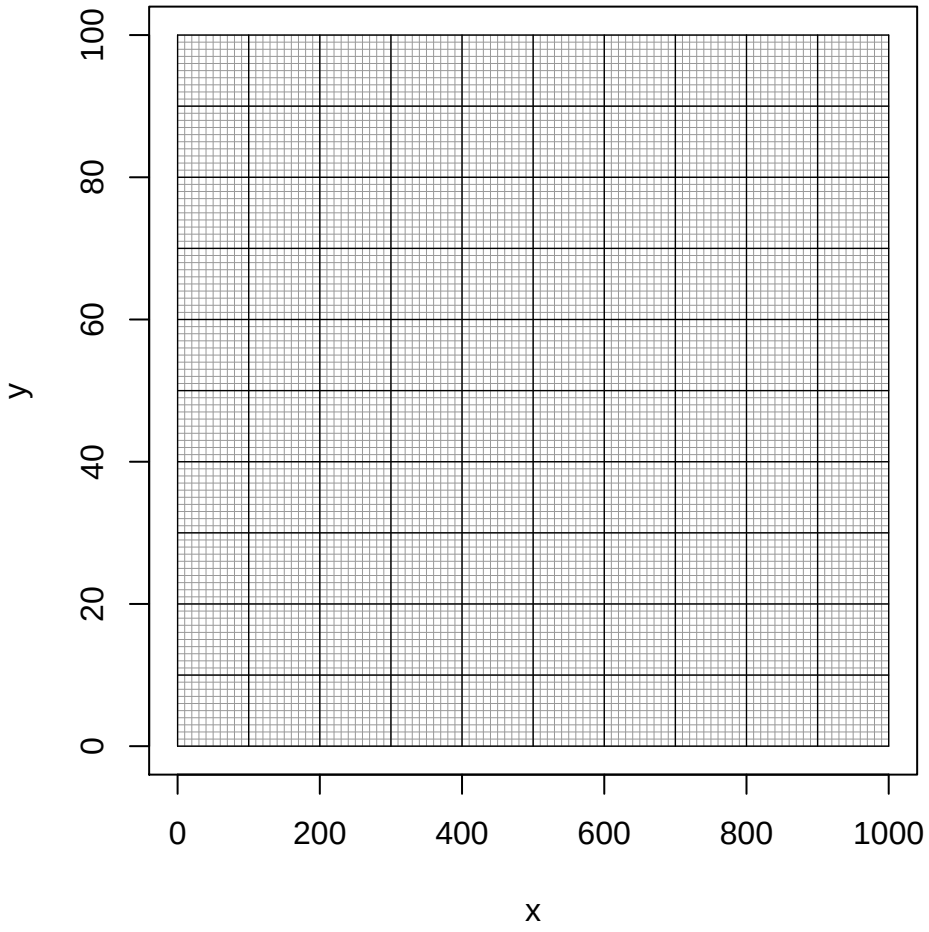
$$b = r\frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

- Write the equation of the regression line (using the calculated values of a and b.)

- On the following page, plot the data and the regression line.

Please plot the data and a corresponding regression line.



Solution

Remember the formula for the correlation coefficient.

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

We calculate the necessary values.

x	y	xy
550	53	29150
590	47	27730
640	43	27520
660	40	26400
360	58	20880
350	57	19950
$\sum x = 3150$	$\sum y = 298$	$\sum x_i y_i = 151630$
$\bar{x} = 525$	$\bar{y} = 49.67$	
$s_x = 137.2$	$s_y = 7.474$	

$$r = \frac{151630 - (6)(525)(49.67)}{(6-1)(137.2)(7.474)} = -0.942$$

If you didn't round any of the steps up to here, you'd get an exact value which is pretty close to our value.

$$r_{\text{exact}} = -0.9398867$$

The regression line has the form

$$y = a + bx$$

So, a is the y-intercept and b is the slope. We have formulas to determine them:

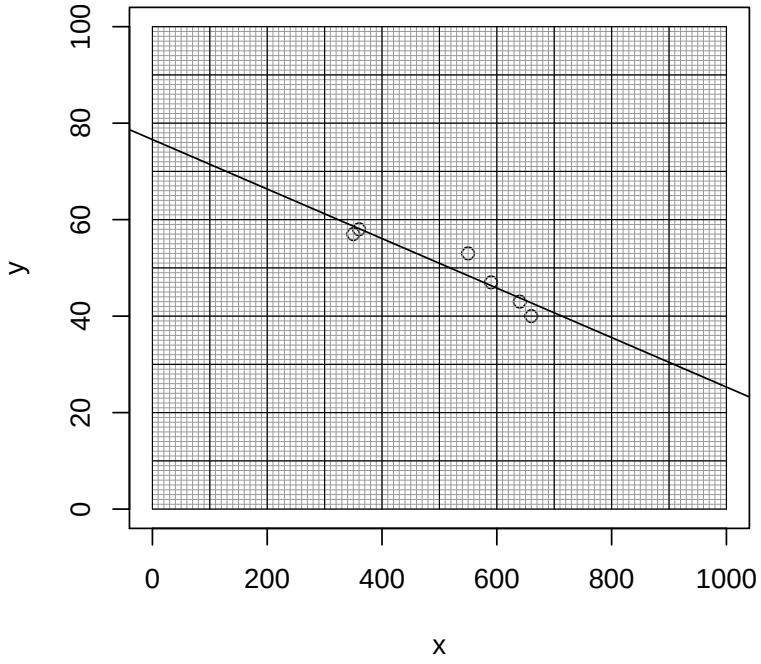
$$b = r\frac{s_y}{s_x} = -0.942 \cdot \frac{7.474}{137.2} = -0.0513$$

$$a = \bar{y} - b\bar{x} = 49.7 - (-0.0513)(525) = 76.6$$

Our regression line:

$$y = 76.6 + (-0.0513)x$$

Make a plot.



Meta-information

extype: num exsolution: 0 exname: regression extol: 0.01