

1. A random sample of 45 students are each scored for
x = hw1 raw score

y = exam1 raw score

finding (fictitious numbers)

$$\bar{x} = 19$$

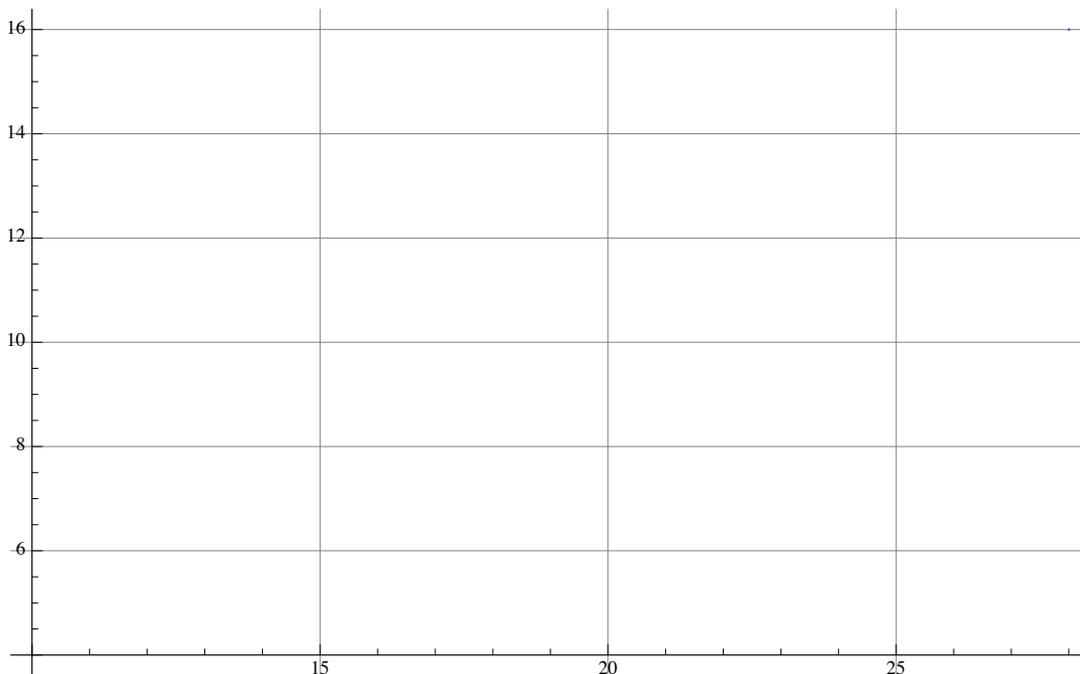
$$\bar{y} = 10$$

$$s_x = 3$$

$$s_y = 2$$

$$r = 0.8$$

1a. Plot the least squares (i.e. regression) line, showing the point of means, sample standard deviations, and correlation r as recognizable elements in your plot. Also plot the naive line.



1b. In your plot 1a overlay a sketch of an elliptical plot whose regression is plotten in 1a.

1c. Sketch bell curves for the distributions of x and y respectively if 1a (elliptical plot) applies.

1d.

2. Make a correction to Lecture 2-6-09b, last page, where the calculation of r is faulty. Verify the correct calculation below using

a.
$$r = \frac{\sum_{i=1}^3 z_{x,i} z_{y,i}}{n-1}$$

b.
$$r = \frac{\overline{xy} - \bar{x} \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \sqrt{\overline{y^2} - \bar{y}^2}}$$

c. your calculator

My error happened when elements of (a) and (b) were inappropriately mixed together (rushing a bit at the close of class, sorry).

$$\mathbf{x} = \{0, 0, 3\}$$

$$\mathbf{y} = \{6, 3, 0\}$$

$$\text{Mean}[\mathbf{x} \mathbf{y}] - \text{Mean}[\mathbf{x}] \text{Mean}[\mathbf{y}] = -3$$

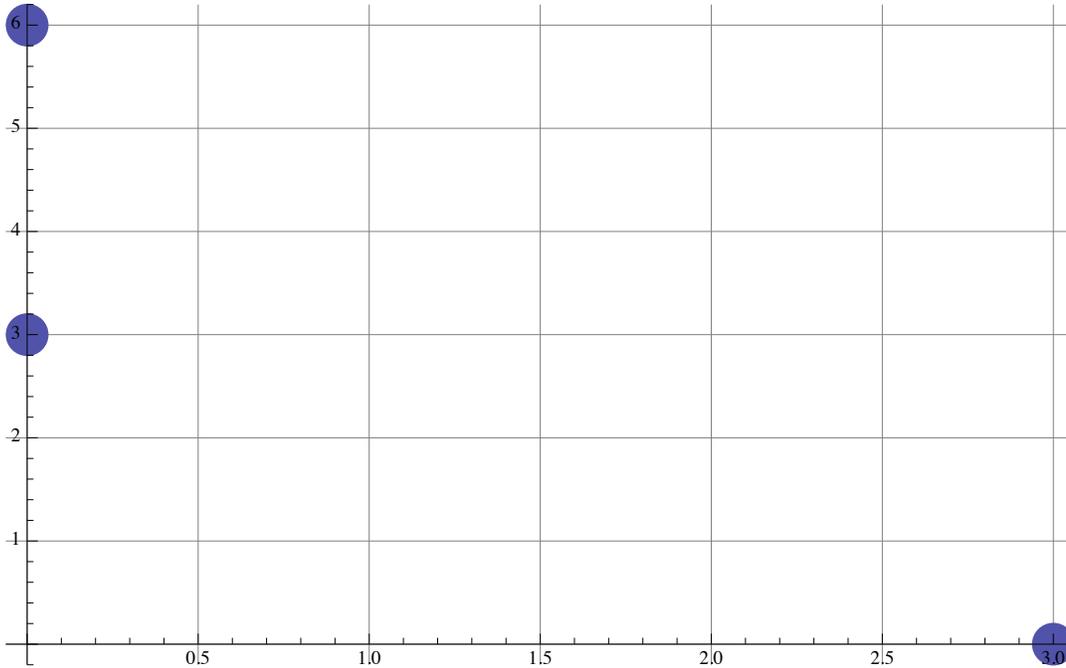
$$\text{Sqrt}[\text{Mean}[\mathbf{x}^2] - \text{Mean}[\mathbf{x}]^2] = \sqrt{2}$$

$$\text{Sqrt}[\text{Mean}[\mathbf{y}^2] - \text{Mean}[\mathbf{y}]^2] = \sqrt{6}$$

$$-3 / \text{Sqrt}[12] = -\frac{\sqrt{3}}{2}$$

$$\mathbf{r} = \frac{\overline{\mathbf{x}\mathbf{y}} - \bar{\mathbf{x}}\bar{\mathbf{y}}}{\sqrt{\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2} \sqrt{\overline{\mathbf{y}^2} - \bar{\mathbf{y}}^2}} = \frac{0 - 1(3)}{\sqrt{3-1^2} \sqrt{15-3^2}} = -\frac{\sqrt{3}}{2}$$

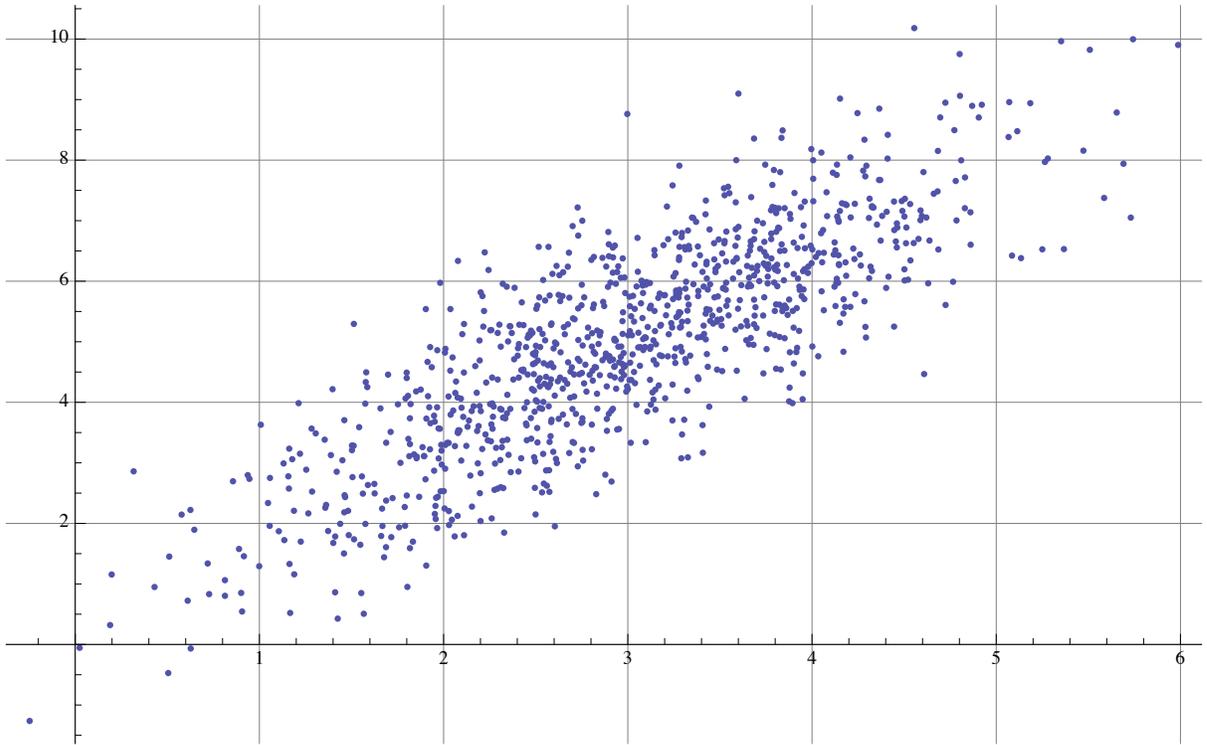
d. Use your calculations (2) to plot the least squares (i.e. regression) and naive lines.

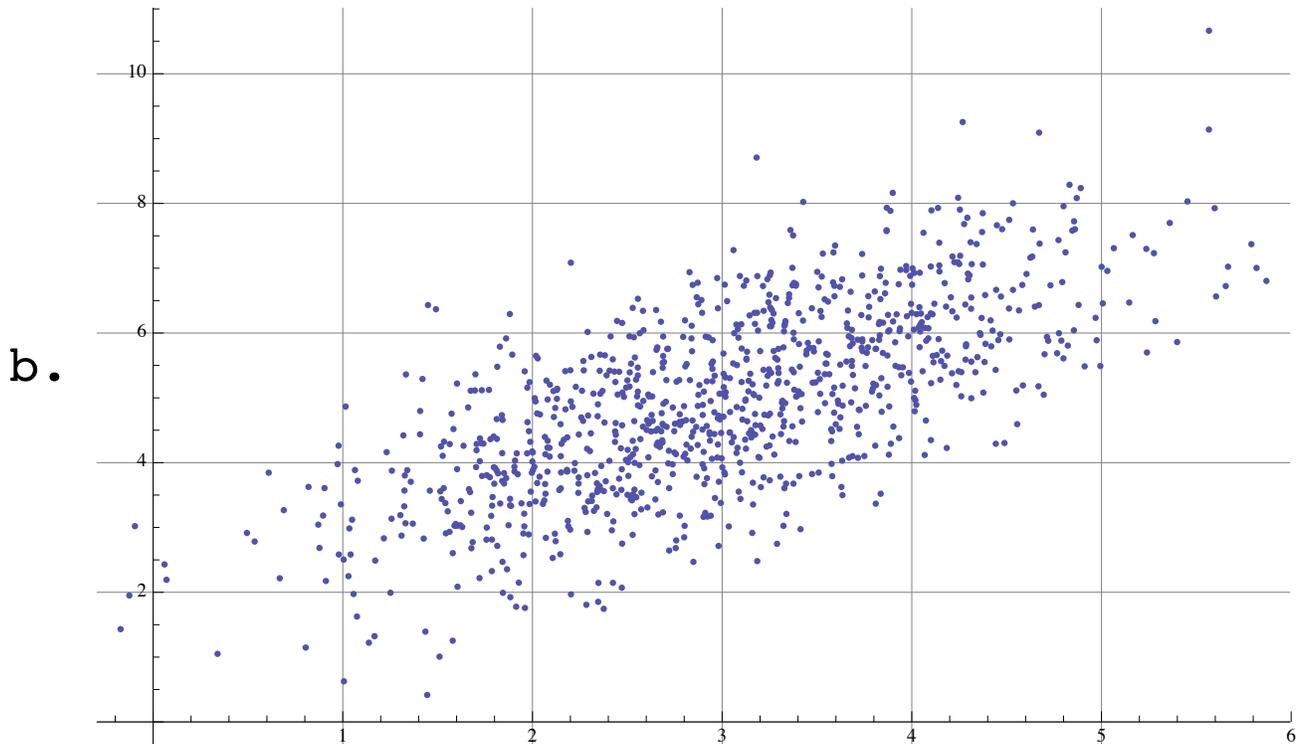


e. This three point plot is not elliptical. Nonetheless, verify the the regression line joins the plot of vertical strip averages. For elliptical plots this is always true, but it is not always the case for non-elliptical plots.

3. For the plots below determine the regression line by eye (as best you can). Read off (as best you can) the means of x , y , the standard deviations of x , y (remember, around 68% of the plot is withing plus or minus one standard deviation of the mean, for each of x and y), and the correlation r .

a.





4. A random sample of 100 students are each scored for
 $x = \text{hw1 raw score}$
 $y = \text{exam1 raw score}$
 finding (fictitious numbers)

$$\bar{x} = 19$$

$$\bar{y} = 10$$

$$s_x = 3$$

$$s_y = 2$$

$$r = 0.8$$

Assume that the plot is elliptical (i.e. 2-D normal).

- a. A student is one standard deviation above average score on hw1. Predict by how many standard deviations they will exceed the mean on exam 1.

- b. A student scores 22 on hw1. Predict their score on exam 1.

- c. Determine the average exam 1 score of all students scoring 22 on hw1.

- d. Predict the exam1 score of a student scoring 25 on hw1.

- e. Predict the exam1 score of a student scoring 14 on hw1.

- f. Illustrate the concept of "regression to mediocrity?"