Solutions to Recitation 10-6-09

These activities cover the essential material from chapters 7, 8.

This coming Monday 5th there will be a graded assignment for you to complete in class for credit.It will be just like your recitation but for different data. I will describe the recitation assignment.Attached you will find a.pdf file of an (x, y) plot. It is the one you will use for your recitation assignment.

** ** *** NOTE : In lecture Monday you will pick up a page with your picture and another plot of this same kind. You will have the opportunity to work in small groups and ask questions about answering the following questions. You will hand in our completed sheet at the end of class FOR CREDIT. You may lose points if you fail to complete this assignment or do poorly on it. ** ** ***

Back to the Tuesday assignment.

For Tuesday recitation Oct. 6 you are to you are to identify the following in the plot attached to this message.

1. Marginal normal density for x. All dots projected on to the x-axis.

2. Standard deviation of marginal distribution of x. eyeball: distance from the center of the normal distr. to the point of inflexion $\sigma_x = 3$

3. Marginal density for y.
   All dots projected on to y-axis.

4. Standard deviation of marginal distribution for y. $\sigma_y = 5$

5. 68% interval for x. interval : $\mu_x - \sigma_x$ to $\mu_x + \sigma_x$

6. 68% interval for y. interval : $[\mu_y - \sigma_y, \mu_y + \sigma_y]$

   diff. notation for same thing

7. Naive line and its slope (rise to run).

To distinguish this line from other lines, check its slope.

slope of naive line $= \dfrac{\sigma_y}{\sigma_x} = \dfrac{5}{3}$

Check: if this value (above) $= \dfrac{\text{rise}}{\text{run}}$ obtained from plot

8. Regression line (the line joining points (x, mean y score for this x). These are the vertical strip means.

9. Slope of regression line. slope of regression line = $r\frac{\sigma_y}{\sigma_x}$ (but in this case we can find it by looking at $\frac{\text{rise}}{\text{run}}$)

10. Ratio (slope of regression line)/(slope of naive line). This ratio is FOR NORMAL PLOTS the correlation r.

$$r = \frac{\text{slope of reg. line}}{\text{slope of naive line}} = \frac{1}{\frac{5}{3}} = \frac{3}{5} = 0.6$$

Also solve the following.

11. Plot points (0, 0), (0, 4), (4, 4). Draw in the "best line" according to you. Is the plot normal?

12. Refer to #11. Calculate mean x, mean y, s(x), s(y), correlation r.

x y
0 0
0 4
4 4

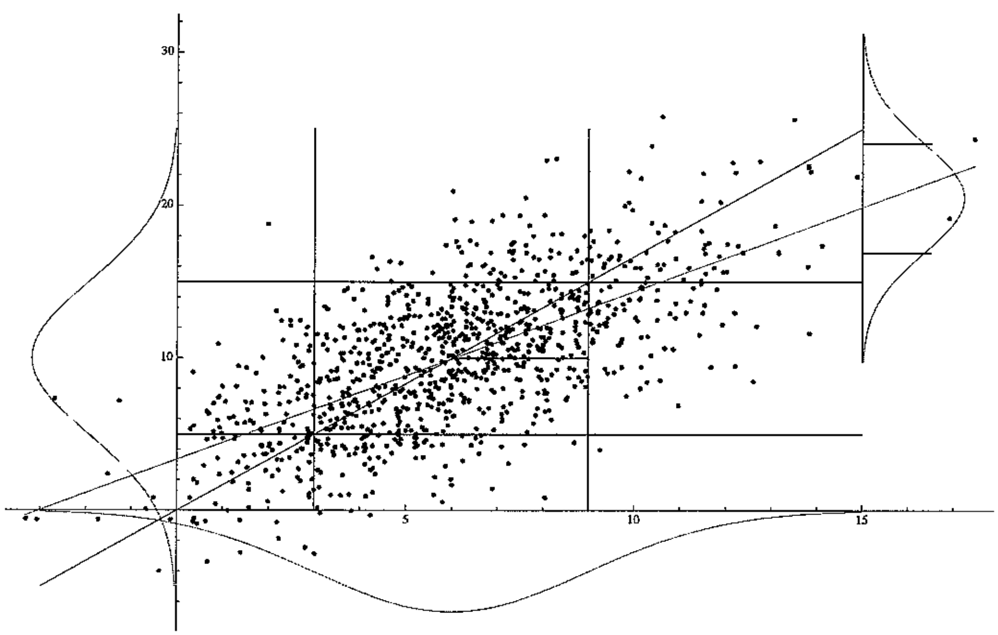
13. Refer to #12. The LEAST SQUARES LINE is defined as that line having the property that it minimizes its distance from the plot of points, in the sense of sum of squares of vertical discrepancies between point and line. Mathematically, the regression line passes through the point of means (mean of x, mean of y) with slope r s(y)/s(x). Find the least squares line for the data of #12 and plot it with that data. Is it the line you chose in #11?

Un-related to the above.

14. The correlation between (x, y) is 0.78. What is the correlation between (2 x - 6, 0.7 y + 3)?

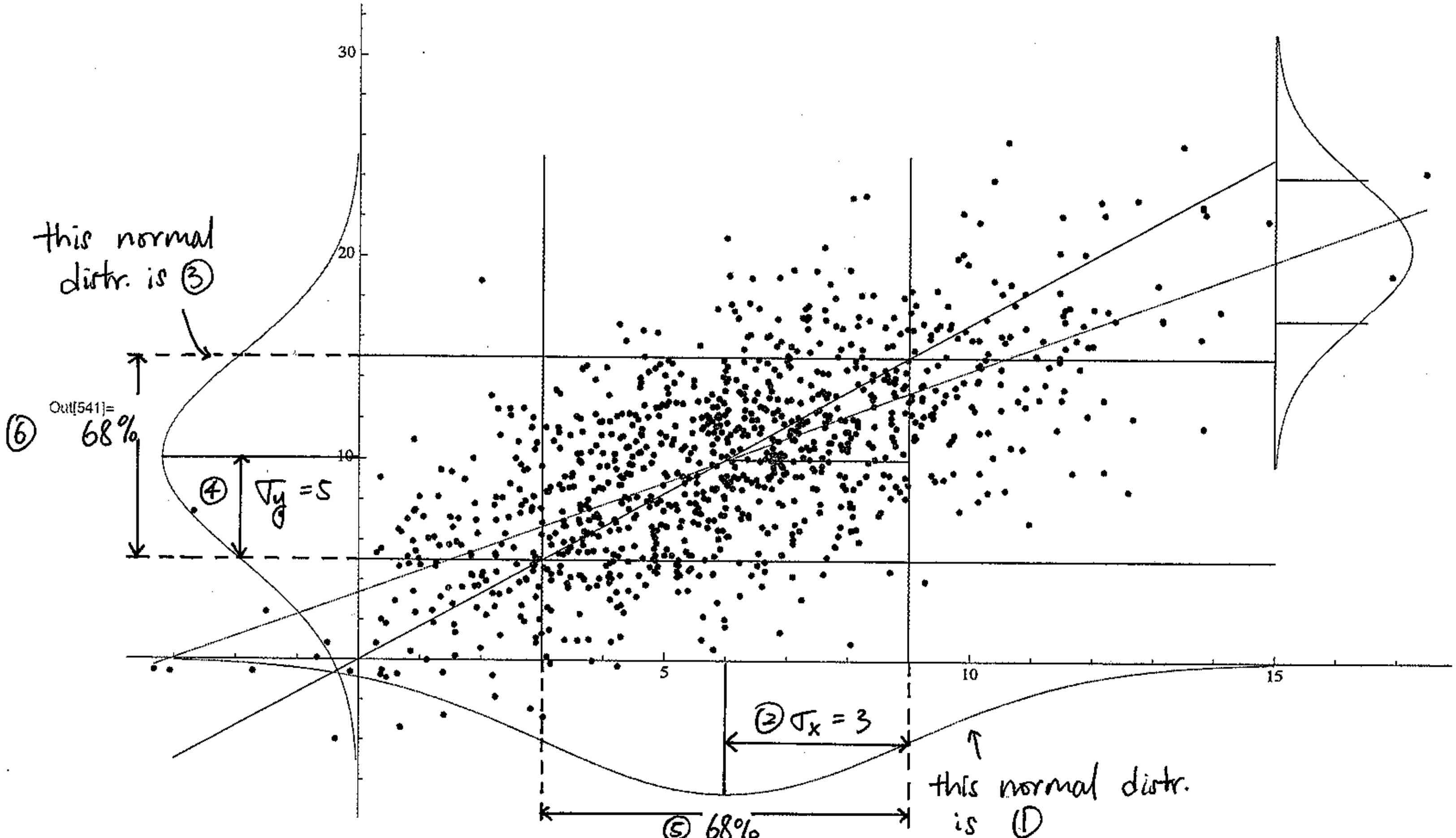
15. The correlation between (x, y) is 0.78. What fraction of s(y) is explained by regression on x?

Out[541]=



Fraction of s_y explained by regression, location of s_y, s_x in picture,
sd of y - scores for given x (algebraically and as seen in picture),
slopes of regr line (algebraically and as seen in picture),
r (algebraically and as seen in picture),
plot of vertical strip averages?

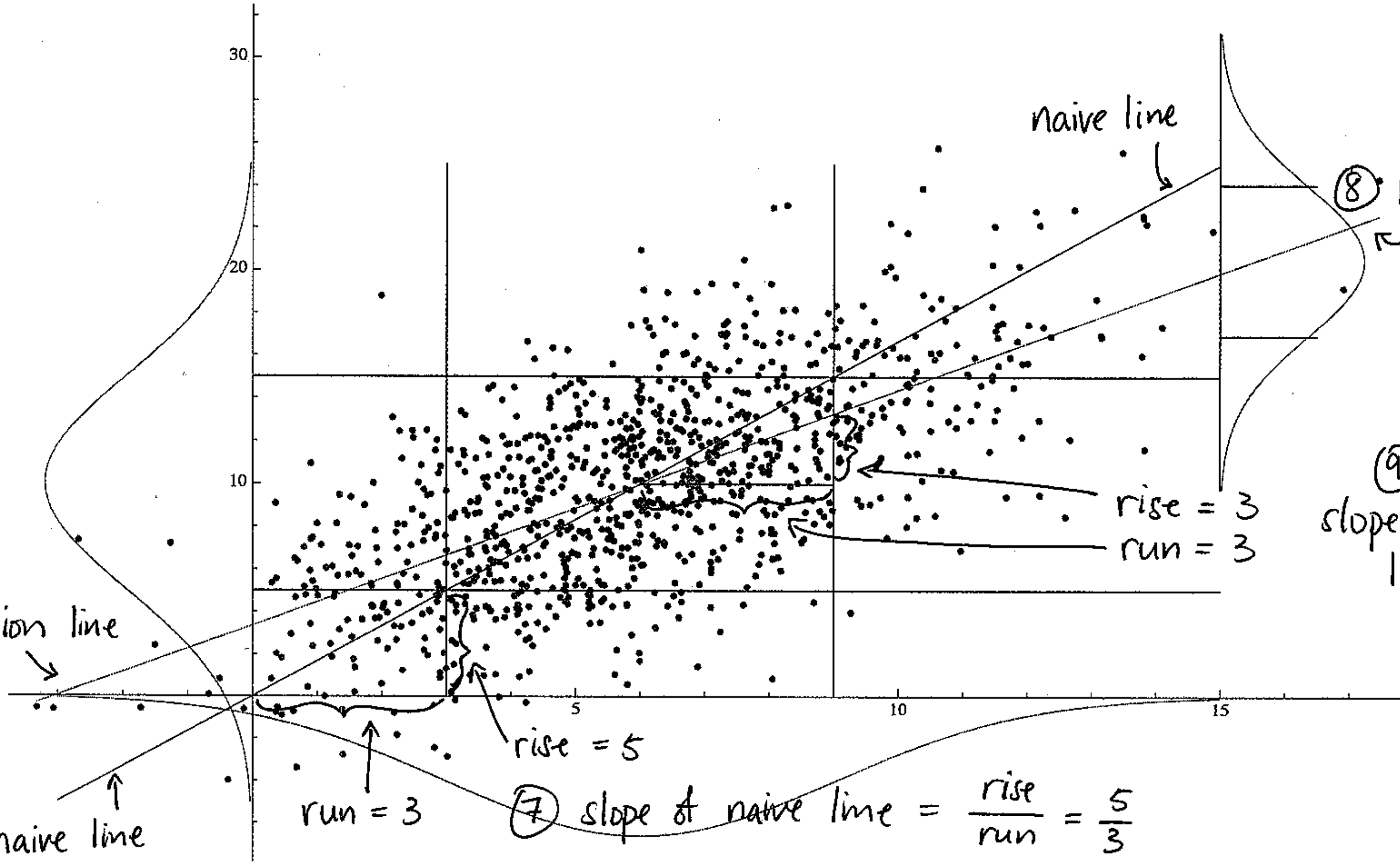
Q. 1-6



Fraction of s_y explained by regression, location of s_y, s_x in picture,
sd of y - scores for given x (algebraically and as seen in picture),
slopes of regr line (algebraically and as seen in picture),
r (algebraically and as seen in picture),
plot of vertical strip averages?

Q. 7 - 9

Out[541]=



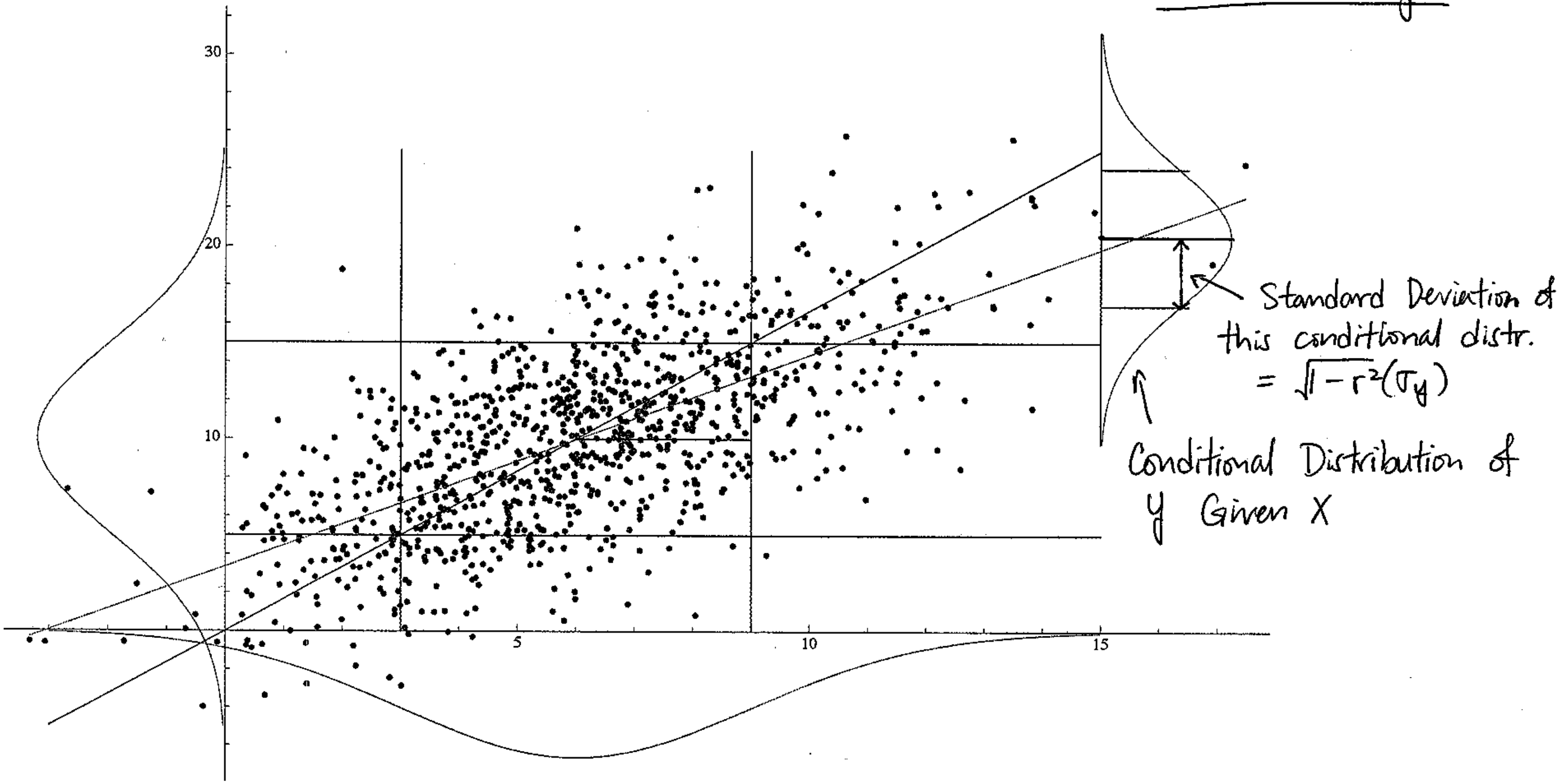
(7) slope of naive line $= \frac{\text{rise}}{\text{run}} = \frac{5}{3}$

(9) slope of regression line $= \frac{3}{3} = 1$

Fraction of s_y explained by regression, location of s_y, s_x in picture,
sd of y - scores for given x (algebraically and as seen in picture),
slopes of regr line (algebraically and as seen in picture),
r (algebraically and as seen in picture),
plot of vertical strip averages?

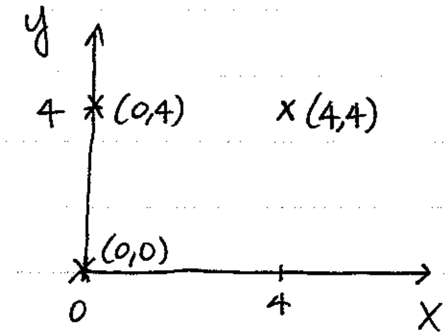
Additional Things

Out[541]=



Fraction of s_y explained by regression, location of s_y, s_x in picture,
sd of y - scores for given x (algebraically and as seen in picture),
slopes of regr line (algebraically and as seen in picture),
r (algebraically and as seen in picture),
plot of vertical strip averages?

11.



The "best line" according to you is just the line of best fit that you think works here. We will actually calculate this line in Q.12–13.

12. Find μ_x, μ_y, σ_x, σ_y, and r.

	x	y	x^2	y^2	$x \cdot y$
	0	0	0	0	0
	0	4	0	16	0
	4	4	16	16	16
Averages	$\bar{x}$	$\bar{y}$			
	μ_x	μ_y	$\overline{x^2}$	$\overline{y^2}$	$\overline{x \cdot y}$
	$4/3$	$8/3$	$16/3$	$32/3$	$16/3$

$$\mu_x = 4/3$$
$$\mu_y = 8/3$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{16/3 - (4/3)^2} = 1.886$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{32/3 - (8/3)^2} = 1.886$$

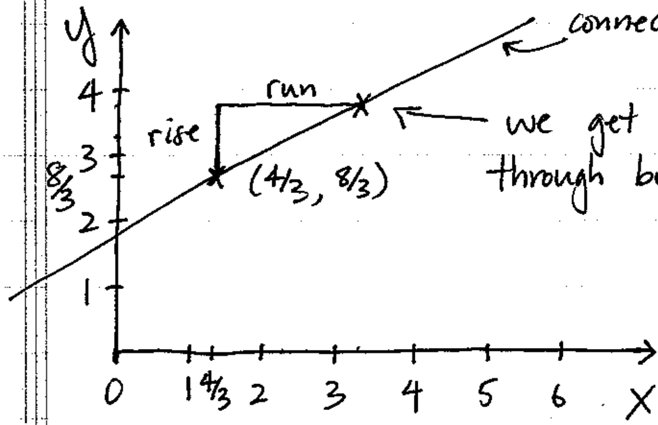
$$r = \frac{\overline{x \cdot y} - (\bar{x})(\bar{y})}{\sigma_x \sigma_y} = \frac{16/3 - (4/3)(8/3)}{(1.886)(1.886)} = 0.50$$

13. To draw the regression line, we need a point that it passes through, and we need its slope.

- regression line (by definition) passes through the point:

$$(\bar{X}, \bar{Y}) = (4/3, 8/3)$$

- slope $= r\dfrac{\sigma_Y}{\sigma_X} = 0.5\dfrac{1.886}{1.886} = 0.5$



connecting the points gives us the regression line

we get this point that the regression line passes through by $\frac{\text{rise}}{\text{run}} = \frac{1}{2} = 0.5$

14. If we manipulate all the data points by:

- multiply by a constant "a"
- adding a constant "b"

in other words "$aX + b$", the correlation is unaffected.
So changing all X values by: $2X - 6$ and all Y values by: $0.7Y + 3$ does not change the correlation r.

$\therefore\ r = 0.78$

15. $r = 0.78 \Rightarrow r^2 = 0.78^2 = 0.608$

Interpretation: 60.8% of the variation in Y, σ_Y^2, (ie. the fluctuation or ~~[illegible]~~ scatter of Y values) is accounted for by the regression on X.