STT200 2-9-09a

1. A random sample of 45 students are each scored for

x = hw1 raw score

y = exam1 raw score

finding (fictitious numbers)

BIG IDEA

$\mu \equiv \quad \sigma \equiv \quad r_{pop}$

EST BY $\bar{x} \equiv \quad s \equiv \quad r_{SAM}$

SAMPLE

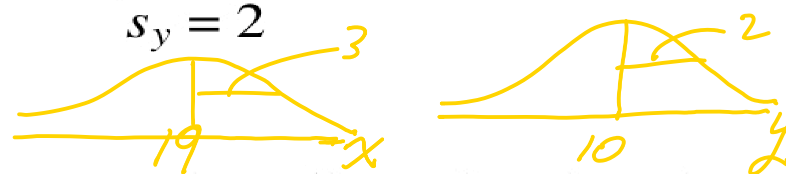
$x_1 \; y_1$

$\vdots \quad \vdots$

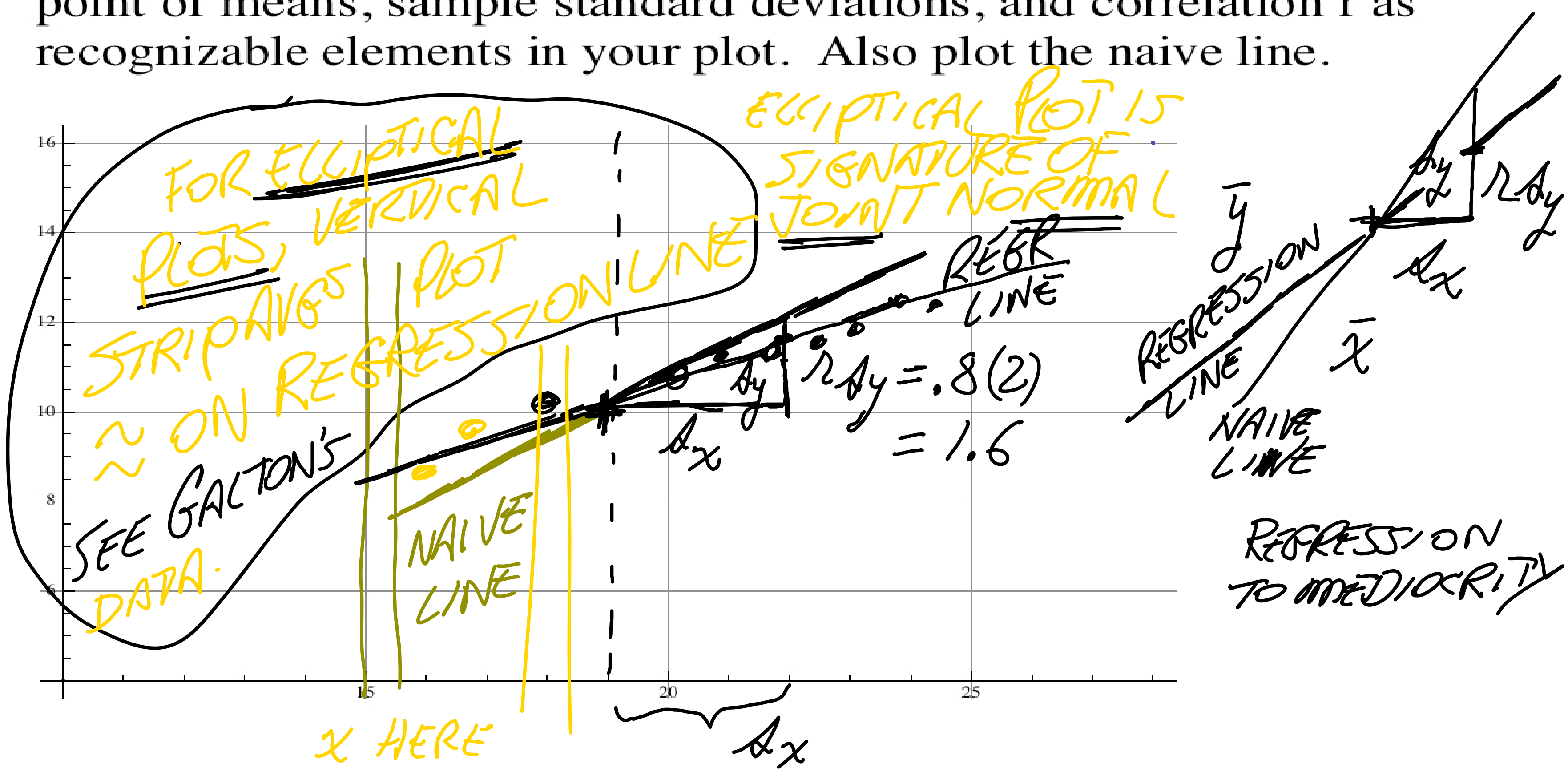
$x_n \; y_n$

$\bar{x} = 19$	$\bar{y} = 10$
$s_x = 3$	$s_y = 2$
r = 0.8	

≈ DIST X



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.



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

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x}$$

1. COND.

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

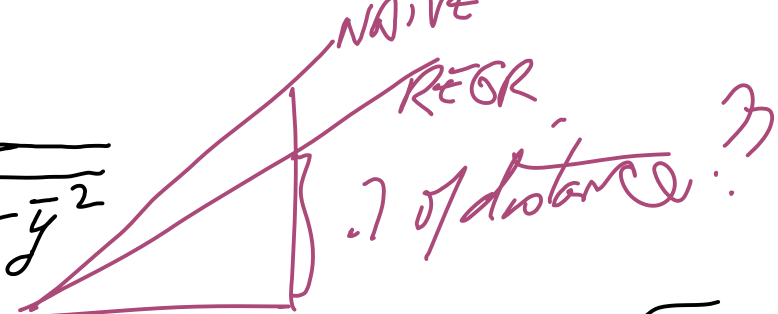
b. $r = \dfrac{\overline{\mathbf{xy}} - \overline{\mathbf{x}}\,\overline{\mathbf{y}}}{\sqrt{\overline{\mathbf{x}^2} - \overline{\mathbf{x}}^2}\,\sqrt{\overline{\mathbf{y}^2} - \overline{\mathbf{y}}^2}}$

CAREFUL! SENSITIVE TO ROUNDING ERRORS.

c. your calculator $\bar{x}$ $\bar{y}$ s_x s_y r

LAY OFF 1 SD EITHER SIDE OF $\bar{x}$?

x = {0, 0, 3}
y = {6, 3, 0} $n = 3$ PAIRS



	x	y	x^2	y^2	xy
	0	6	0	36	0
	0	3	0	9	0
	3	0	9	0	0
AVG	1	3	3	15	$0 = \overline{xy}$

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

$$= \frac{0 - 1(3)}{\sqrt{3 - 1^2}\sqrt{15 - 3^2}} = \frac{-3}{\sqrt{2}\sqrt{6}} = \frac{-\sqrt{3}}{2}$$

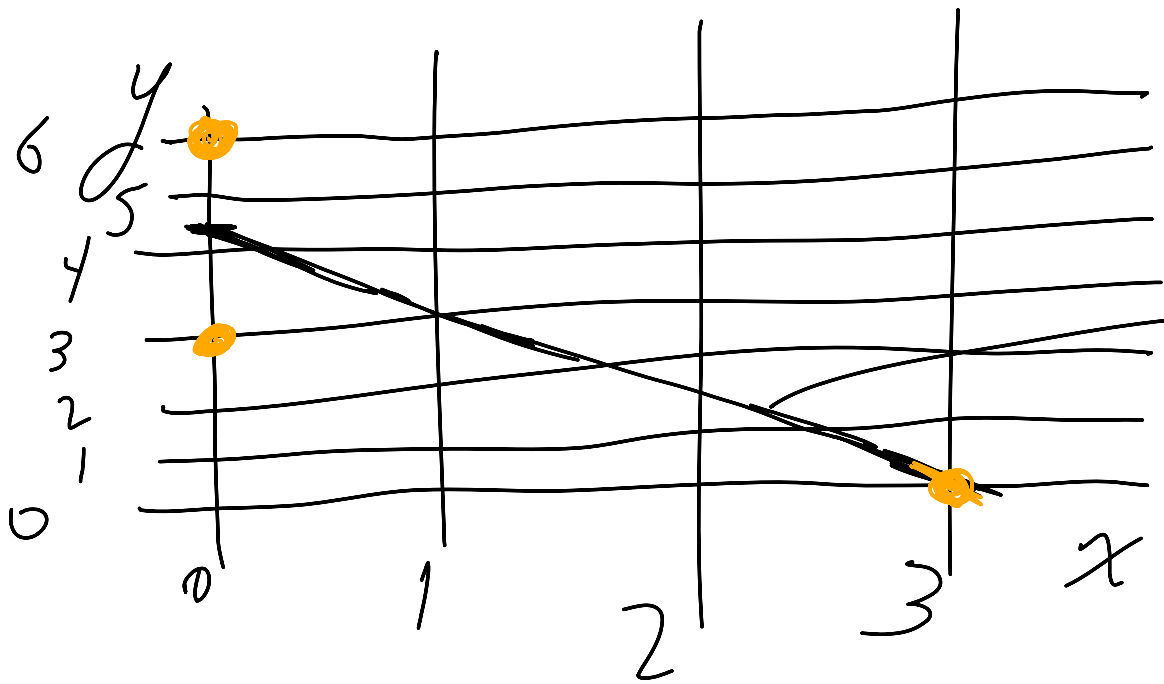
SEE $\bar{x}\bar{y} = 1(3) = 3 \neq \overline{xy} = 0$ $\quad \overline{x^2} = 3$ CORRELATION ?

```
x = {0, 0, 3}
y = {6, 3, 0}
Mean[x y] - Mean[x] Mean[y] = -3
Sqrt[Mean[x^2] - Mean[x]^2] = √2
Sqrt[Mean[y^2] - Mean[y]^2] = √6
-3 / Sqrt[12] = -√3/2
```

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

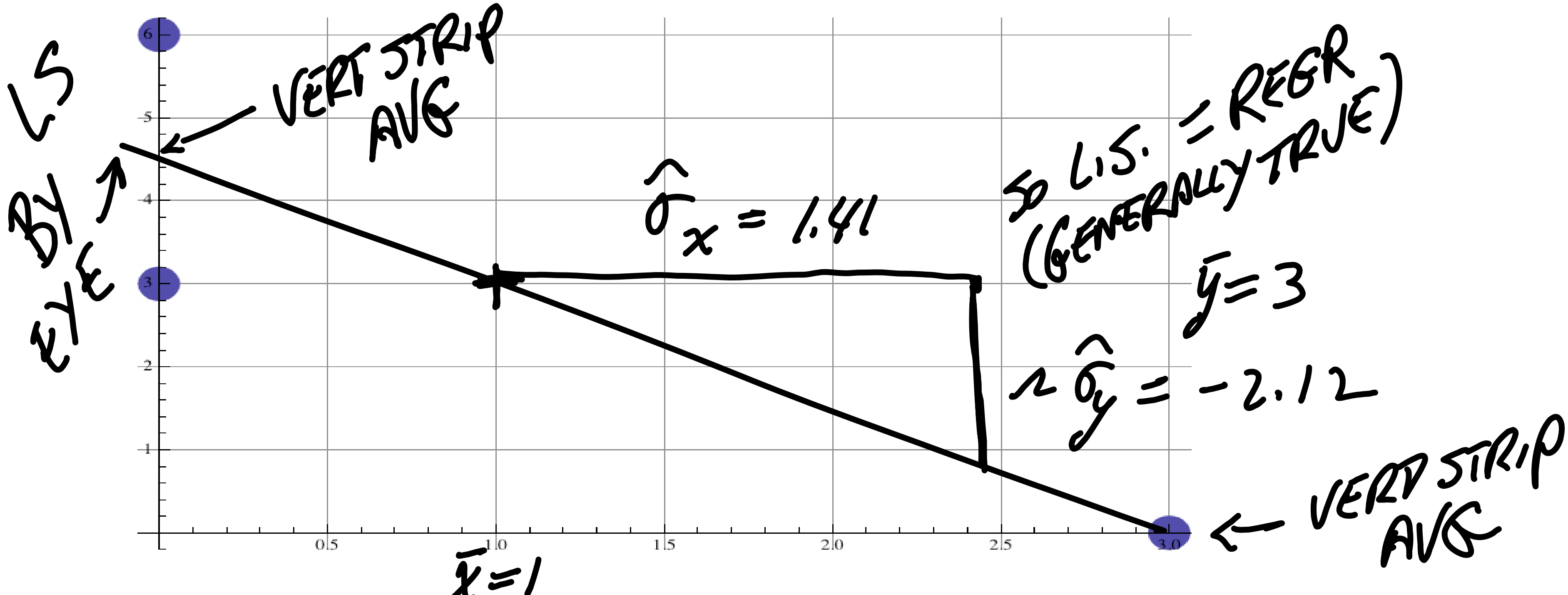
$$r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}} = \frac{0 - 1\,(3)}{\sqrt{3 - 1^2}\,\sqrt{15 - 3^2}} = -\frac{\sqrt{3}}{2}$$

x y
(0, 6)



L.S. LINE

L.S. LINE ALWAYS = REGR LINE



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.

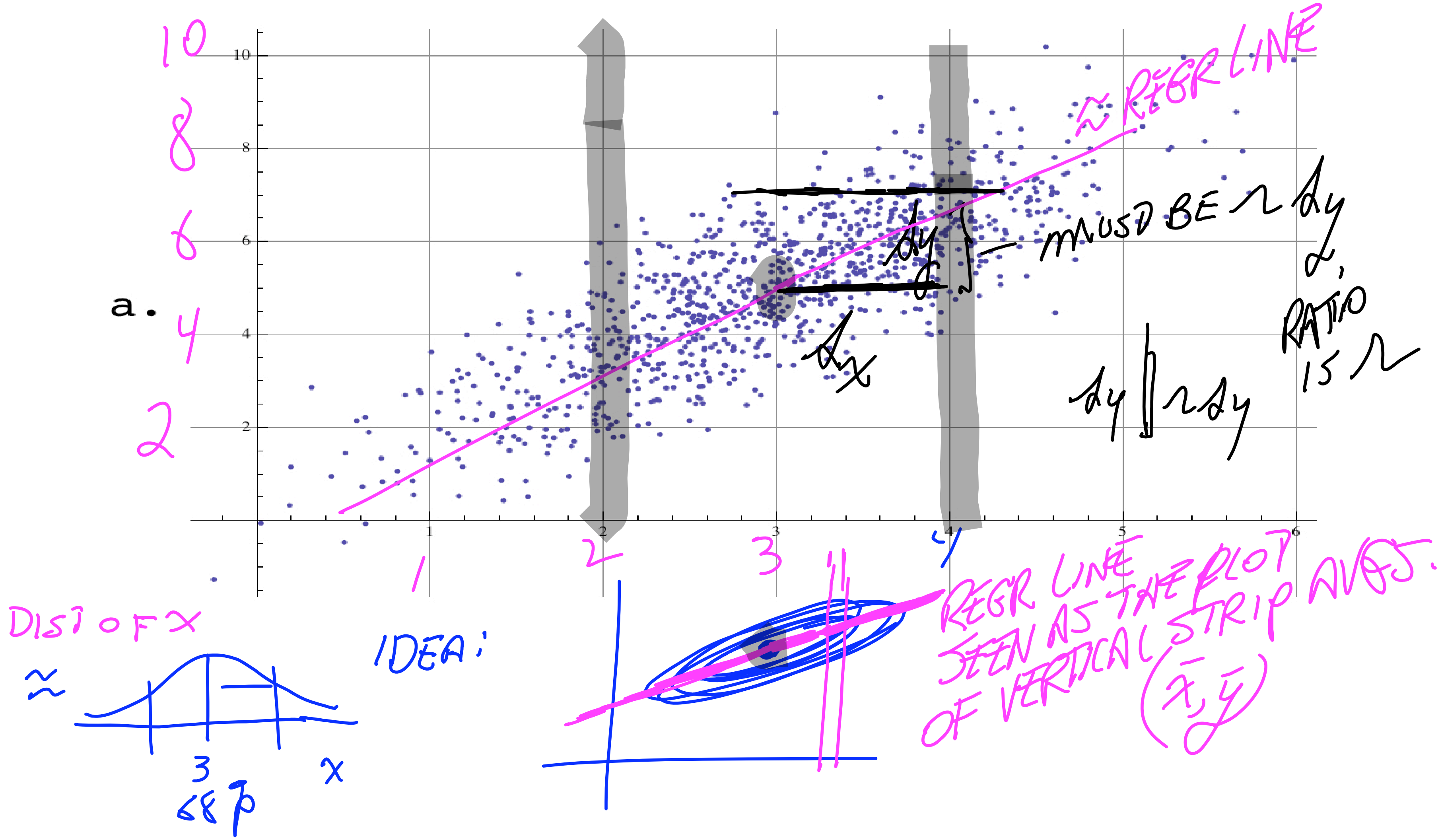
$\hat{\sigma}_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{2}$ 1.41 $\qquad \hat{\sigma}_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{6} = 2.45$

$r = -\sqrt{3}/2 = -0.87 \qquad r\hat{\sigma}_y = -\sqrt{3}\sqrt{6}/2 = -3/\sqrt{2} = -2.12$

(CAN USE INSTEAD $\delta_x, \delta_y, r\delta_y$)

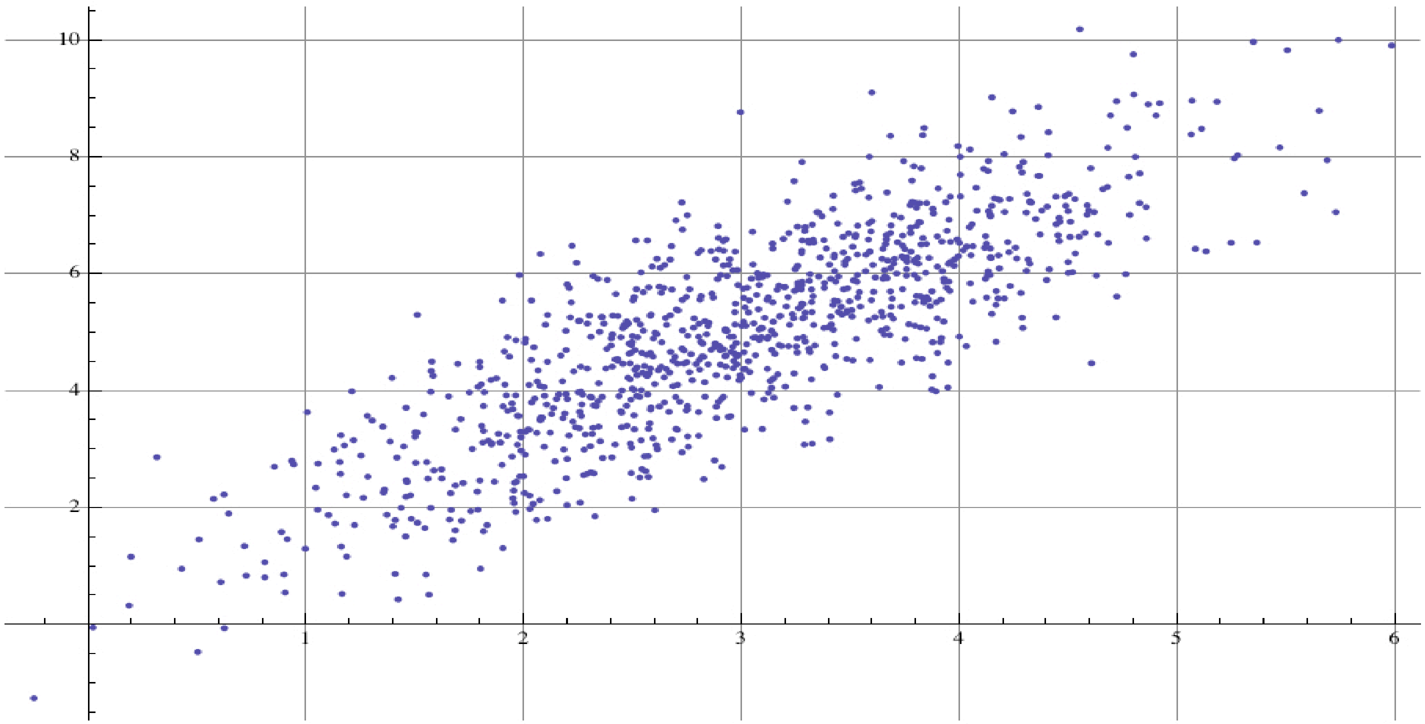
3. For the plots below determine the regression line by eye (as best you can). Read off (as best you can) the means of x, y, the standard deviations of x, y (remember, around 68% of the plot is withing plus or minus one standard deviation of the mean, for each of x and y), and the correlation r.

a.



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 = hw1 raw score

y = exam1 raw score

finding (fictitious numbers)

$\bar{x} = 19$ $\quad\quad$ $\bar{y} = 10$

$s_x = 3$ $\quad\quad$ $s_y = 2$

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.

$\bar{x} + 1 s_x \xrightarrow[\text{ON AVG}]{} \bar{y} + r(1) s_y$ ANS. $0.8\, s_y$

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

$\underset{19+3}{22 = \bar{x} + s_x}$ SO PREDICT $\bar{y} + 0.8 s_y = 10 + 0.8(2) = 11.6$

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

FOR ELLIPTICAL PLOT SEE (b).

STT200
2-9-09b

1. A random sample of 45 students are each scored for

x = hw1 raw score

y = exam1 raw score

finding (fictitious numbers)

$\bar{x}, \bar{y}, s_x, s_y, r$

$\bar{x} = 19$ $\qquad\qquad$ $\bar{y} = 10$

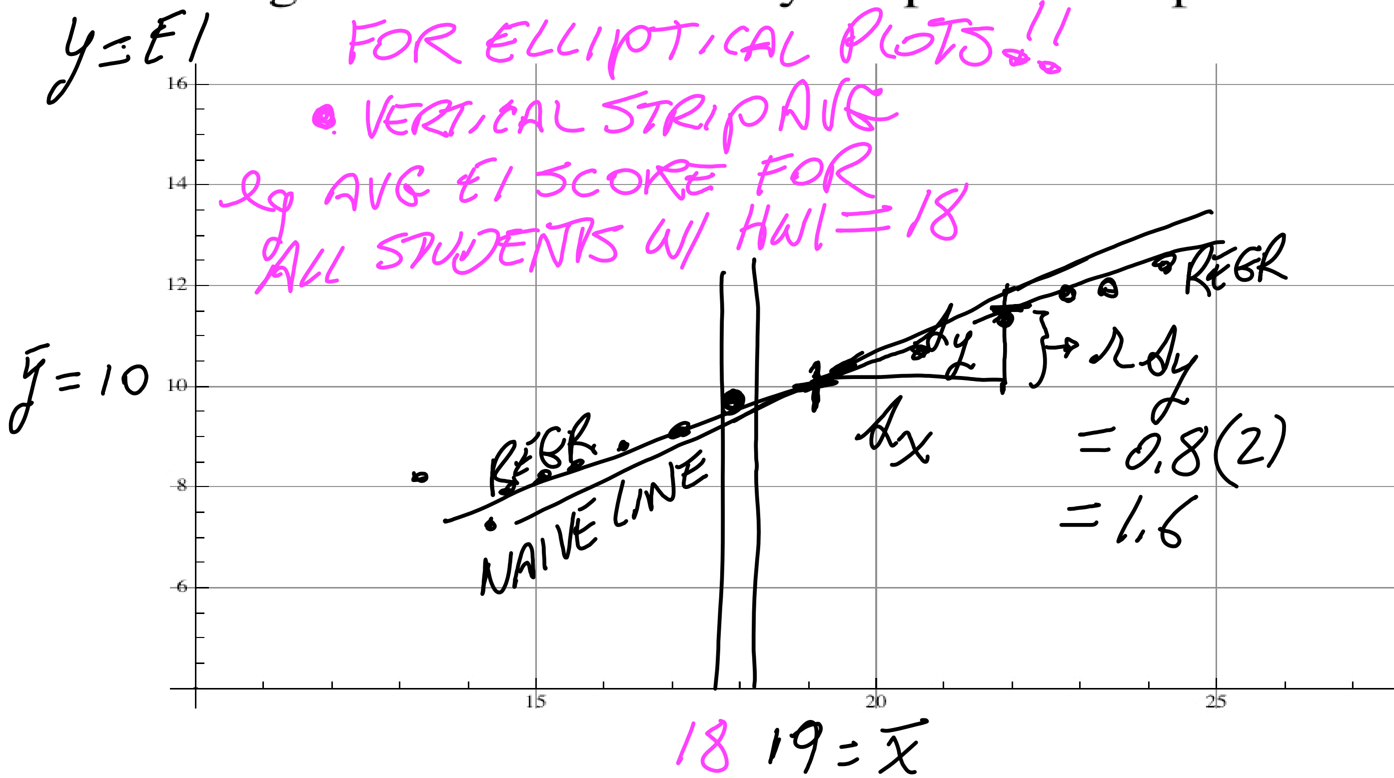
$s_x = 3$ $\qquad\qquad$ $s_y = 2$

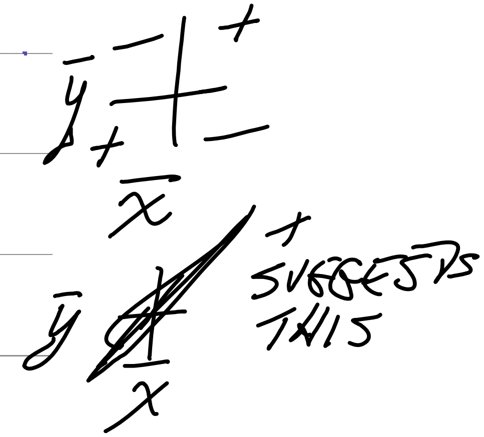
$r = 0.8$

$x, y, w \quad \bar{x}, \bar{y}, \bar{w} \quad s_x\ s_y\ s_w \quad r_{xy}\ r_{xw}\ r_{yw}$

$r[x,y] = r[y,x] = r_{ax+b,\, cy+d} \quad a>0,\ c>0$

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.





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

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x} \qquad z_{y_i} = \frac{y_i - \bar{y}}{s_y}$$

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

SENSITIVE TO ROUNDING

b. $r = \dfrac{\overline{\mathbf{xy}} - \overline{\mathbf{x}}\ \overline{\mathbf{y}}}{\sqrt{\overline{\mathbf{x}^2} - \overline{\mathbf{x}}^2}\ \sqrt{\overline{\mathbf{y}^2} - \overline{\mathbf{y}}^2}} = \dfrac{0 - 1(3)}{\sqrt{3-1^2}\sqrt{15-3^2}} = \dfrac{-3}{\sqrt{2}\sqrt{6}} = \dfrac{-3}{\sqrt{12}}$

$= \dfrac{-\sqrt{3}}{2} = r$ correlation

c. your calculator

x = {0, 0, 3}
y = {6, 3, 0}

x	y	x^2	y^2	xy
0	6	0	36	0
0	3	0	9	0
3	0	9	0	0
AVG 1	3	3	15	0
$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{y^2}$	

Comment: $s_x \neq \sqrt{\overline{x^2} - \bar{x}^2}$

$$s_x = \sqrt{\frac{n}{n-1}}\ \underbrace{\sqrt{\overline{x^2} - \bar{x}^2}}_{\text{ACTUALLY}}$$

WE'LL USE THIS $\hat{\sigma} = \sqrt{\dfrac{(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n}}$

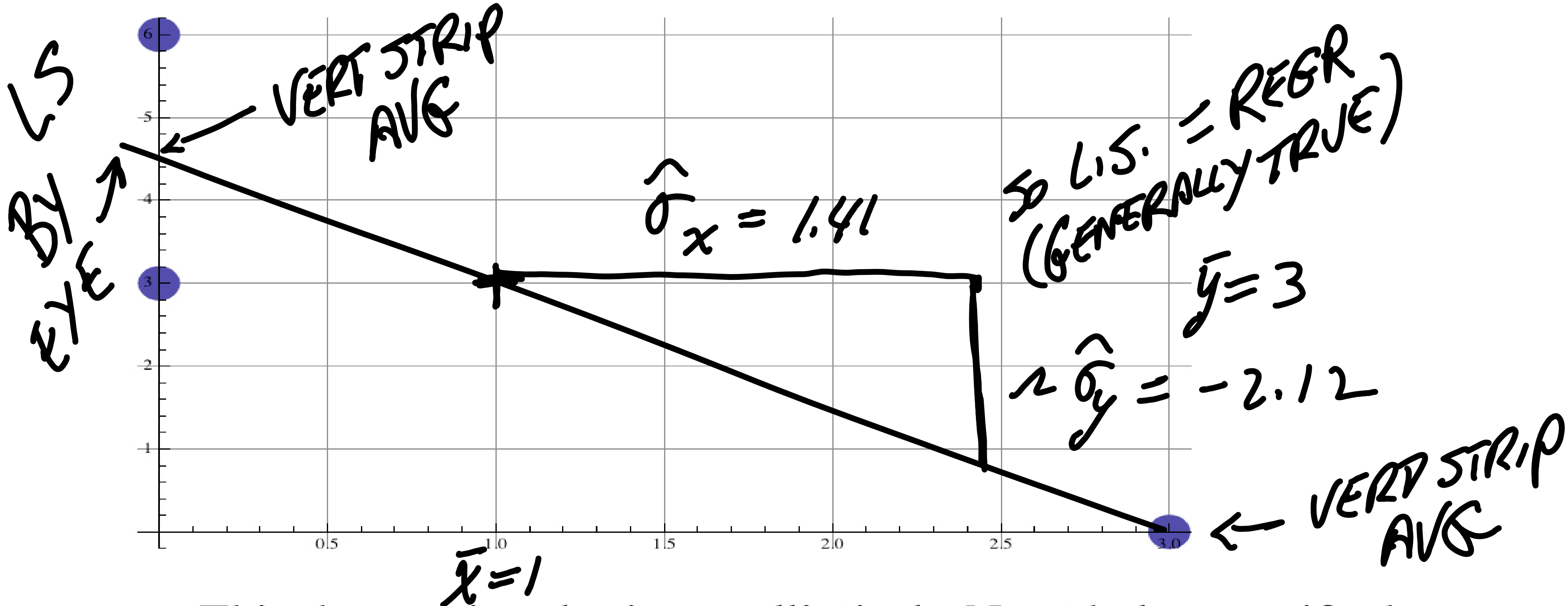
```
x = {0, 0, 3}
y = {6, 3, 0}
Mean[x y] - Mean[x] Mean[y]
```
$= -3$

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

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

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

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



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.

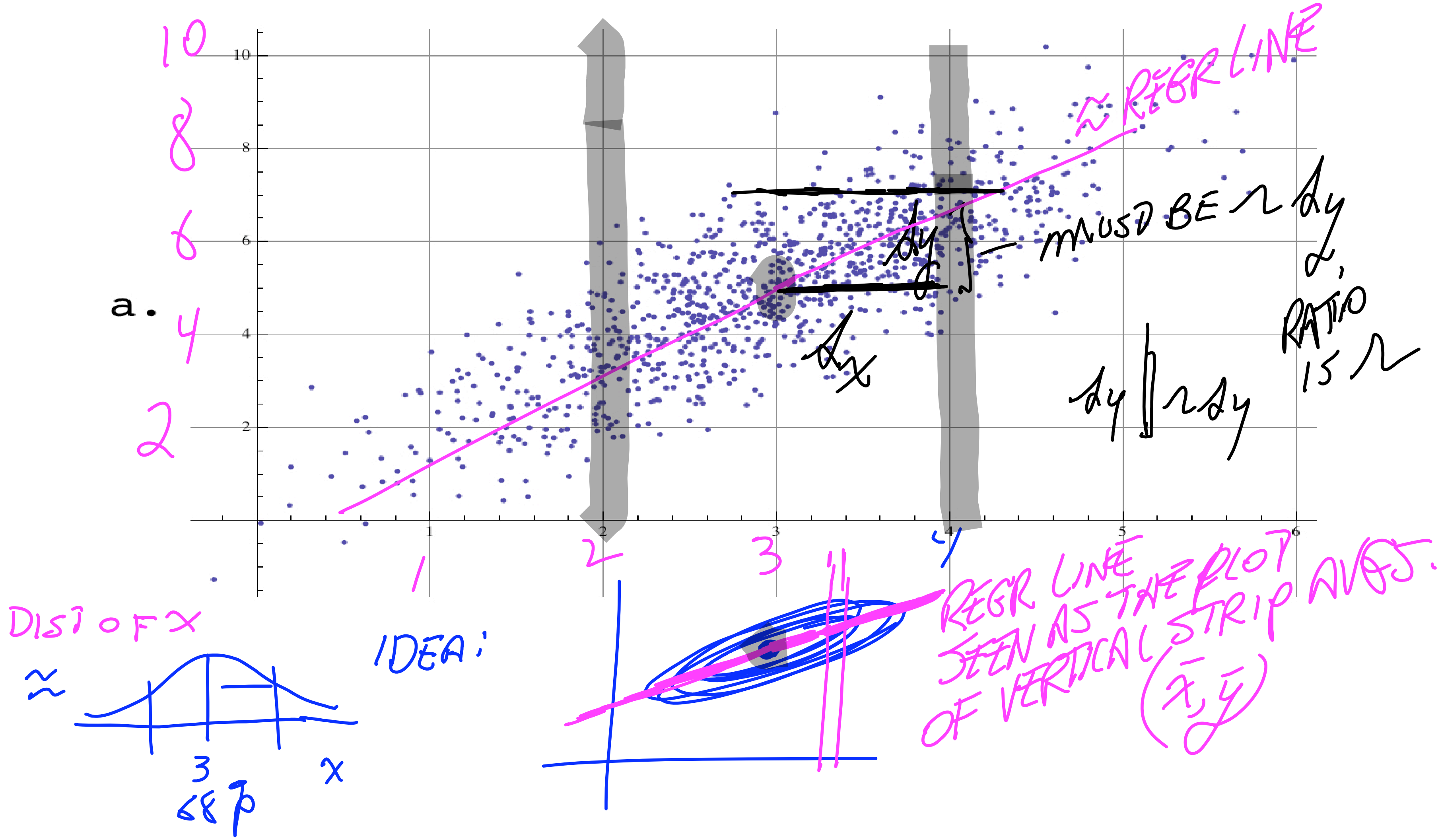
$$\hat{\sigma}_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{2} \quad 1.41 \qquad \hat{\sigma}_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{6} = 2.45$$

$$r = -\sqrt{3}/2 = -0.87 \qquad r\,\hat{\sigma}_y = -\sqrt{3}\sqrt{6}/2 = -3/\sqrt{2} = -2.12$$

CAN USE σ_x, σ_y, $r\sigma_y$ INSTEAD.

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 = hw1 raw score

y = exam1 raw score

finding (fictitious numbers)

$\overline{x} = 19$ $\overline{y} = 10$

$s_x = 3$ $s_y = 2$

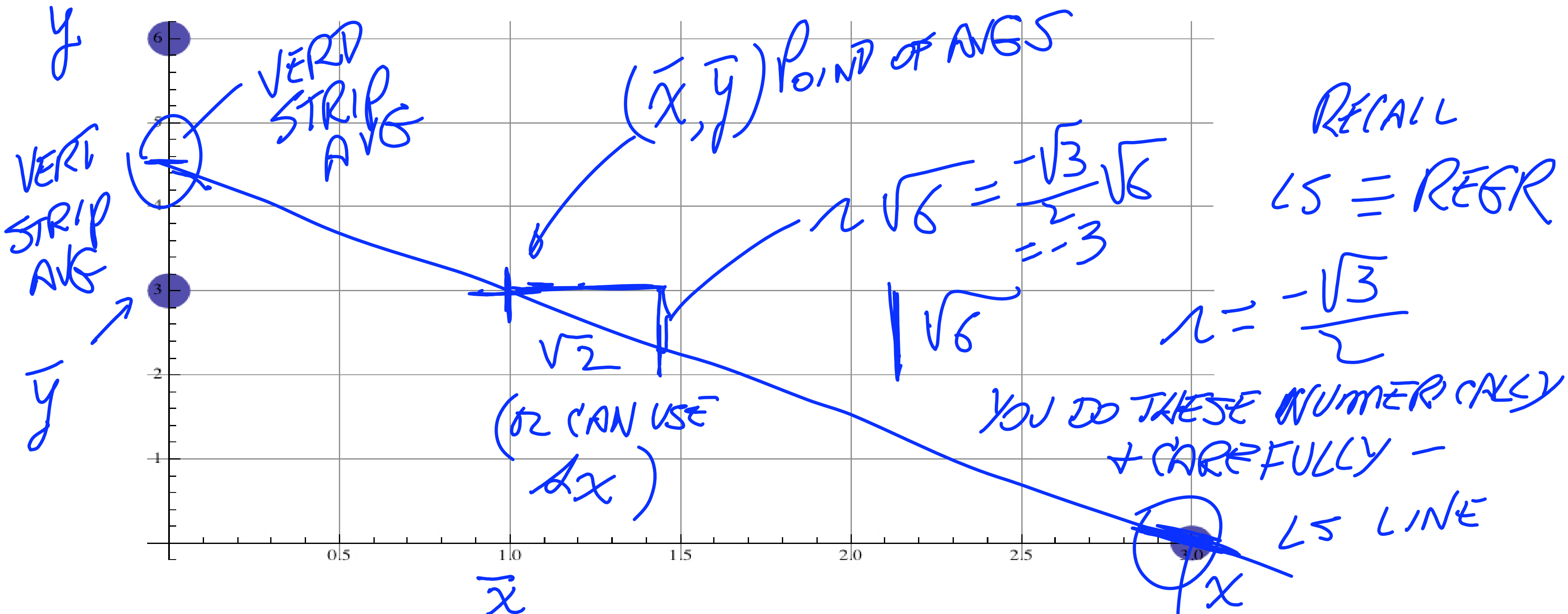
r = 0.8

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.



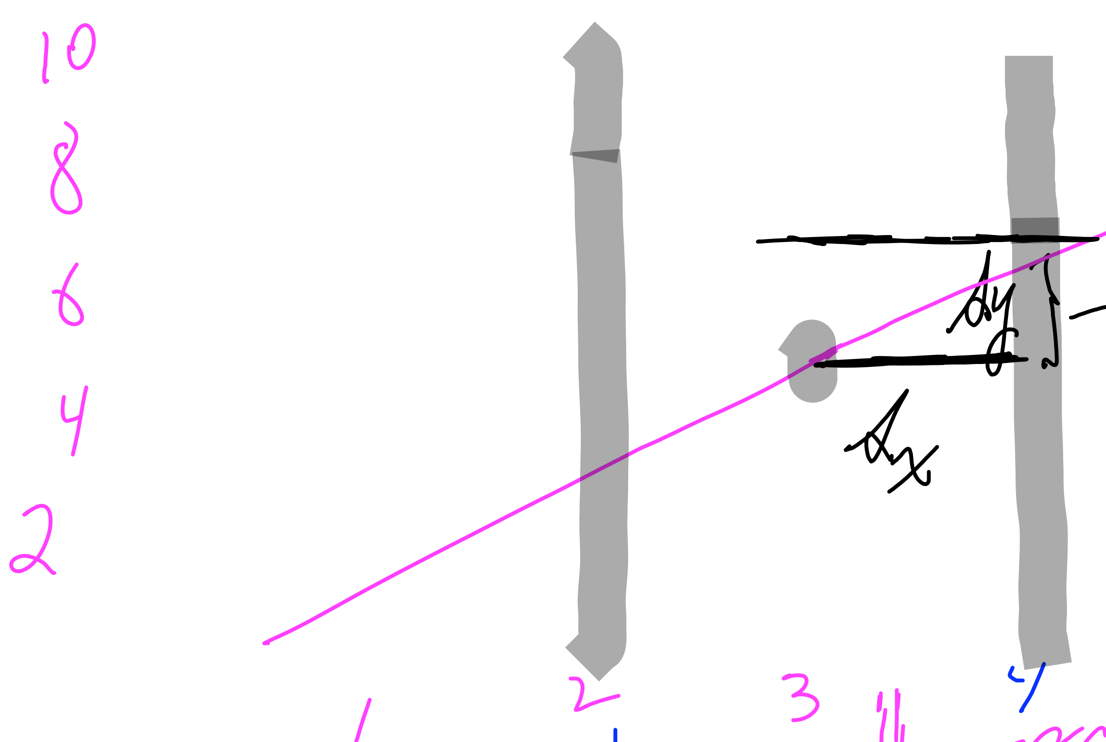
e. This three point plot is not elliptical. Nonetheless, verify the the regression line joins the plot of vertical strip averages. For elliptical plots this is always true, but it is not always the case for non-elliptical plots.

VERT STRIP AVG

(x, y): $(0,6)$, $(0,3)$, $(3,0)$ $n=3$

NOTE: $\underbrace{\sum (y_i - c)^2} = \underbrace{(c - \bar{y})^2}_{\text{PENALTY}} + \underbrace{\sum (y_i - \bar{y})^2}$

APPLY TO VERT STRIP



MUST BE ≈ d_y

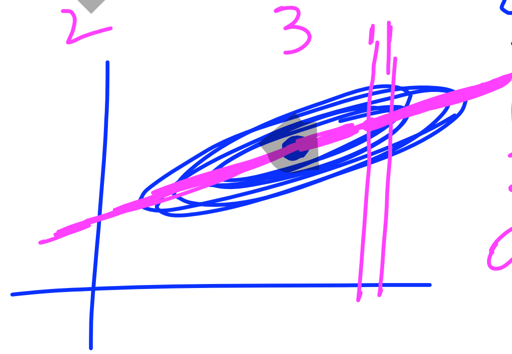
RATIO IS ≈

$d_y \parallel \approx d_y$

DIST OF x

≈

IDEA:



REGR LINE SEEN AS THE PLOT OF VERTICAL STRIP AVGS.

$(\bar{x}, \bar{y})$