$\bar{x}$ $\bar{y}$ $\overline{x^2}$ $\overline{y^2}$ $\overline{xy}$

STP200

2-16-09a

34.01 456. 1194.58 208788. 15391.9

1. s_y. $= \cancel{\sqrt{\frac{n}{n-1}}} \sqrt{208\cancel{7}789 - 456^2}$ 2. r.

n LARGE

$n = \#$ PAIRS

3. The fraction of s_y^2 accounted for by regression on x.

4. The slope of the naive line.

5. The slope of the regression line of y on x.

6. The slope of the regression line of x on y (which would apply if the variables were interchanged).

SPREE $n \sim \infty$

SAMPLE SD OF $x_1 \ldots x_n = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{n}{n-1}} \sqrt{\overline{x^2} - \bar{x}^2}$

SAMPLE CORRELATION =

$\sigma_x = \sqrt{1194 - 34^2}$

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}} = \frac{15392 - (34)456}{\sqrt{1194 - 34^2}\ \sqrt{208789 - 456^2}}$$

time $\bar{x}$ cal $\bar{y}$ $\overline{x2}$ $\overline{y2}$ $\overline{xy}$

34.01 456. 1194.58 208788. 15391.9

7. $r[\cancel{2}\,x - \cancel{4}, \cancel{6}\,y + \cancel{2}]$ 2, 6 HAVE SAME SIGN

8. $r[-x + \cancel{2}, y - \cancel{6}] = -r[x, y]$ −1, 1 OPPOSITE SIGNS

9. For an ELLIPTICAL plot having the above averages, the aver age calories for all subjects having time 36. $\approx \bar{y} = 456$

36 CLOSE TO $\bar{x}$

10. For an ELLIPTICAL plot having the above averages, the best (by least squares) prediction for calories for a student with time 36. $\bar{y} + (36 - \bar{x})\, r\, \hat{\sigma}_y / \hat{\sigma}_x$ (YOU WORK IT OUT)

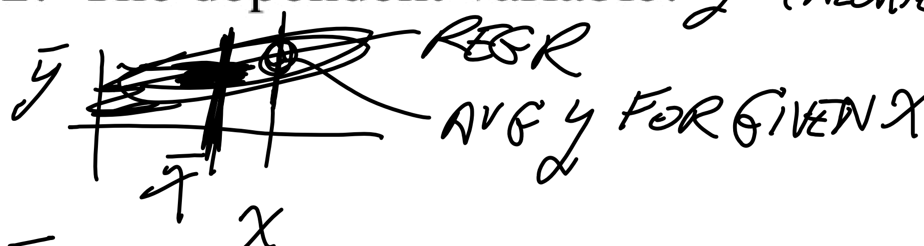
11. The independent variable.

ALWAYS X (TIME)

$\dfrac{y - \bar{y}}{x - \bar{x}} = r\, \hat{\sigma}_y / \hat{\sigma}_x$ SLOPE

12. The dependent variable. y CALORIES

#9.



$\bar{x} = 34$ $\bar{y} = 456$

13-21. r[x, y] = 0.9, $s_x = 2$, $s_y = 5$, $\overline{x} = 22$, $\overline{y} = 54$.

13. Determine r[y, x]. $= r[x, y]$ SAME !!

14. For points (x, y) on the regression line determine the numerical value of $\frac{y - \overline{y}}{x - \overline{x}}$. $=$ SLOPE $= r\, \hat{\sigma}_y / \hat{\sigma}_x = r\, s_y / s_x$ DIVISORS n or n-1 CANCEL

15. For x = $\overline{x}$ + s_x the regression prediction of y is $\overline{y}$ + (?) s_y.

$\bar{x} + 2\, s_x \rightarrow$ PRED $\bar{y} + .9(2)\, s_y$ 0.9

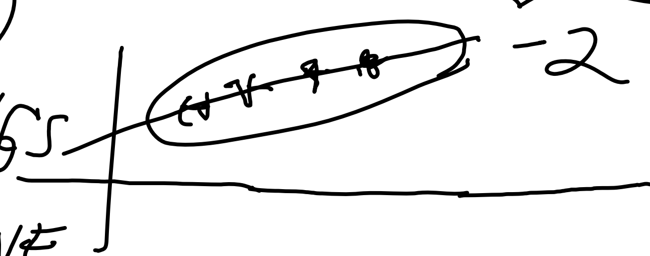
16. For x = 18 the regression prediction of y is?

17. Regression predictions (15), (16) are sometimes useful even if the plot is not elliptical. If the plot IS ELLIPTICAL what is the special nature of the plot of vertical strip averages?

#16. $\bar{y} + r\, s_y$ (STD SCORE OF 18) $= \bar{y} + .9\, s_y \left(\frac{18-22}{2}\right)$ -2

ANS $\bar{y} + .9\, s_y (-2)$

#17. PLOT OF VERT STRIP AVGS ≈ REGRESSION LINE



r[x, y] = 0.9, $s_x = 2$, $s_y = 5$, $\bar{x} = 22$, $\bar{y} = 54$.

18. If the plot is ELLIPTICAL what is the average y-score for all (x, y) pairs with x = 18? PT ON REGR LINE AT $x = 18$

19. If the plot is ELLIPTICAL what is the standard deviation of y-scores for all (x, y) pairs with x = 18?

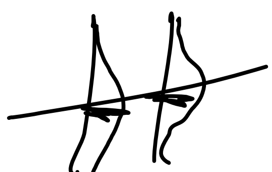
20. If the plot is elliptical, sketch the distribution of x, and the distribution of y.

21. Draw a picture illustrating all of (18), (19), (20).

#18. $$\frac{y - \bar{y}}{x - \bar{x}} = r \frac{s_y}{s_x} = .9\left(\frac{5}{2}\right)$$ SOLVE FOR y

18

SO ANS. AVG AT $x = 18$ IS $\underset{\bar{y}}{54} + (18 - 22).9\left(\frac{5}{2}\right)$

#19.  SD $\sqrt{1 - r^2}\, s_y = \sqrt{1 - .81}\,(5)$

SO WITHOUT X YOU GUESS $y \approx \bar{y}$ AVG. BUT KNOWING $x = 18$ GUESS REGR.

GUESS $\bar{y}$ SD $\approx s_y = 5$

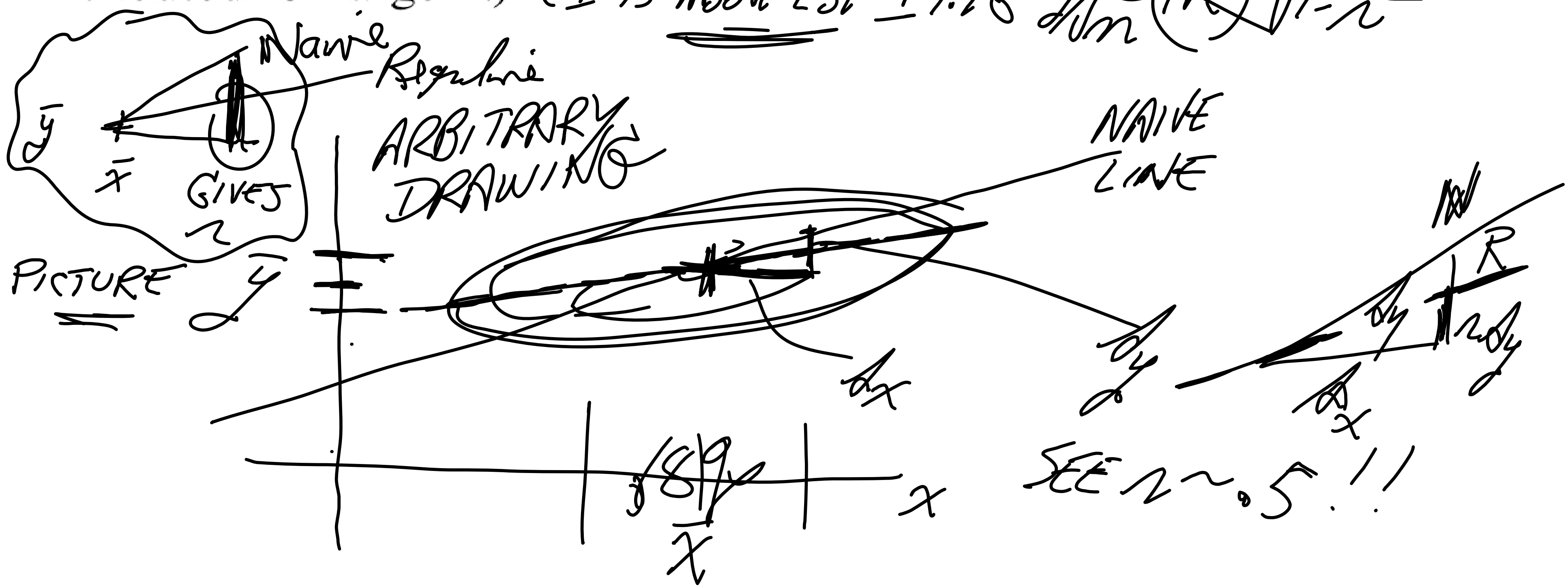
REGR GUESS SD $\approx \sqrt{1 - .9^2}\,(5)$

22-23. r[x, y] = 0.9, s_x = 2, s_y = 5, $\overline{x}$ = 22, $\overline{y}$ = 54.

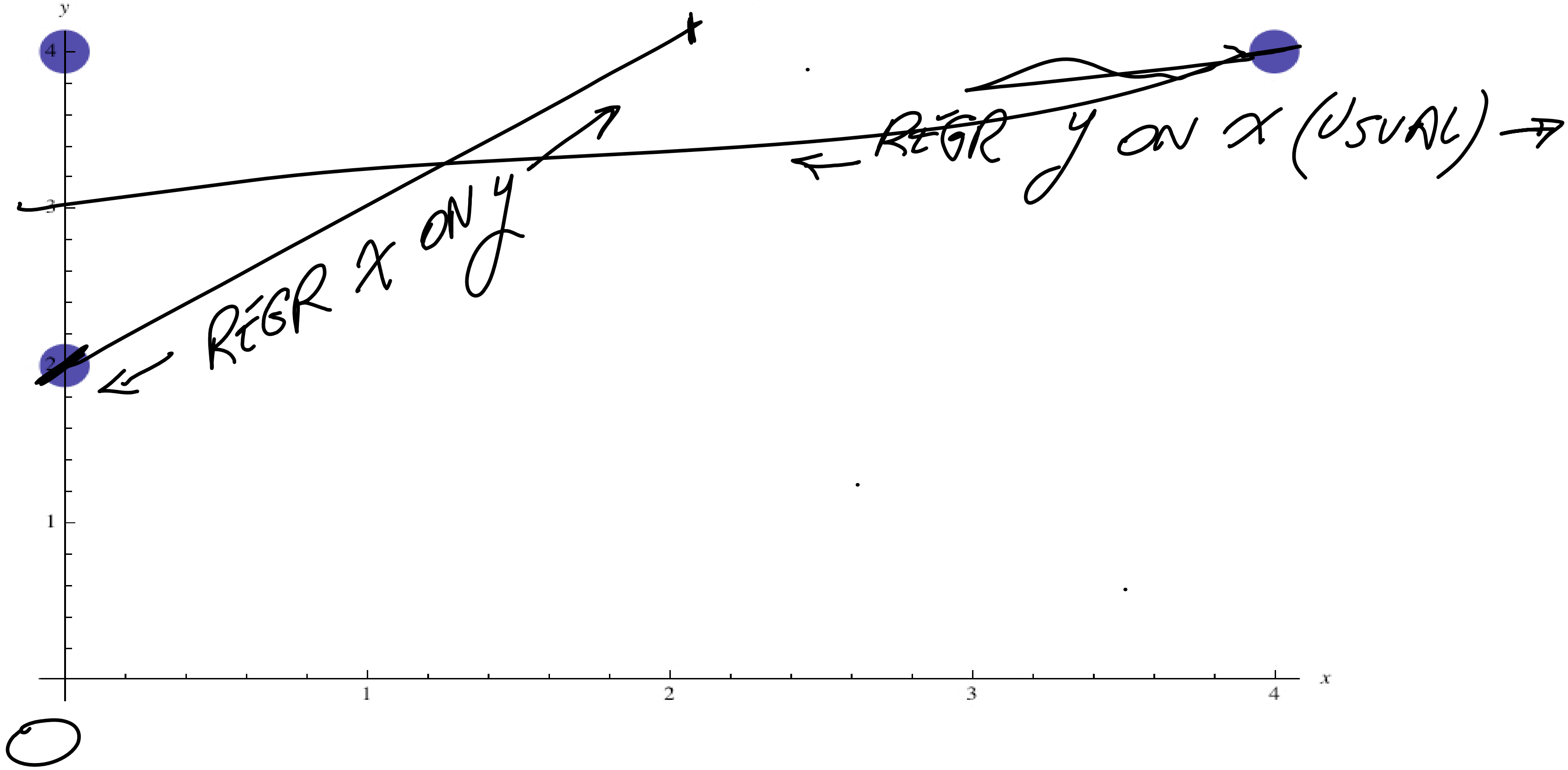
22. Using (14), if I tell you that the population mean of x is μ_x = 26 what is the regression-based estimate for μ_x?

REGR BASED EST OF μ_y INSERT μ_x TO REGR. EST OF $\mu_y = \bar{y} + (\mu_x - \bar{x})\, r\, s_y/s_x$ 26

23. Give the 95% CI for the estimate (19) if n is large. (The plot need not be elliptical since the estimator (19) is dependent upon $\overline{x}$ and $\overline{y}$ which are approximately jointly normal distributed for large n.) CI IS ABOVE EST $\pm 1.96\, s_y/\sqrt{n}\, (FPC) \sqrt{1-r^2}$



24. For the plot below, sketch the regression line for y on x (usual). Also, sketch the regression line for x on y.



= regrtable[{0, 0, 4}, {2, 4, 4}]

atrixForm=

x	y	x^2	y^2	xy
0	2	0	4	0
0	4	0	16	0
4	4	16	16	16
$\overline{1.33333}$	$\overline{3.33333}$	$\overline{5.33333}$	$\overline{12.}$	$\overline{5.33333}$

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}} = \frac{5.33 - 1.33 \; 3.33}{\sqrt{5.33 - 1.33^2}\sqrt{12 - 3.33^2}} = .5$$

25. For the plot just above calculate the slope of regression. Confirm it with what you see in the plot.

$$\text{SLOPE} = r\,\hat{\sigma}_y/\hat{\sigma}_x = \frac{5.33 - 1.33 \; 3.33}{5.33 - 1.33^2}$$

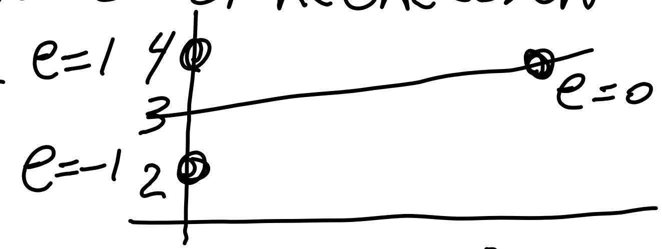
See that in general

26. From your calculations (25) what is s_y?

$$\sqrt{12 - 3.33^2}$$

27. Calculate s_e.

e ARE THE RESIDUALS OF REGRESSION



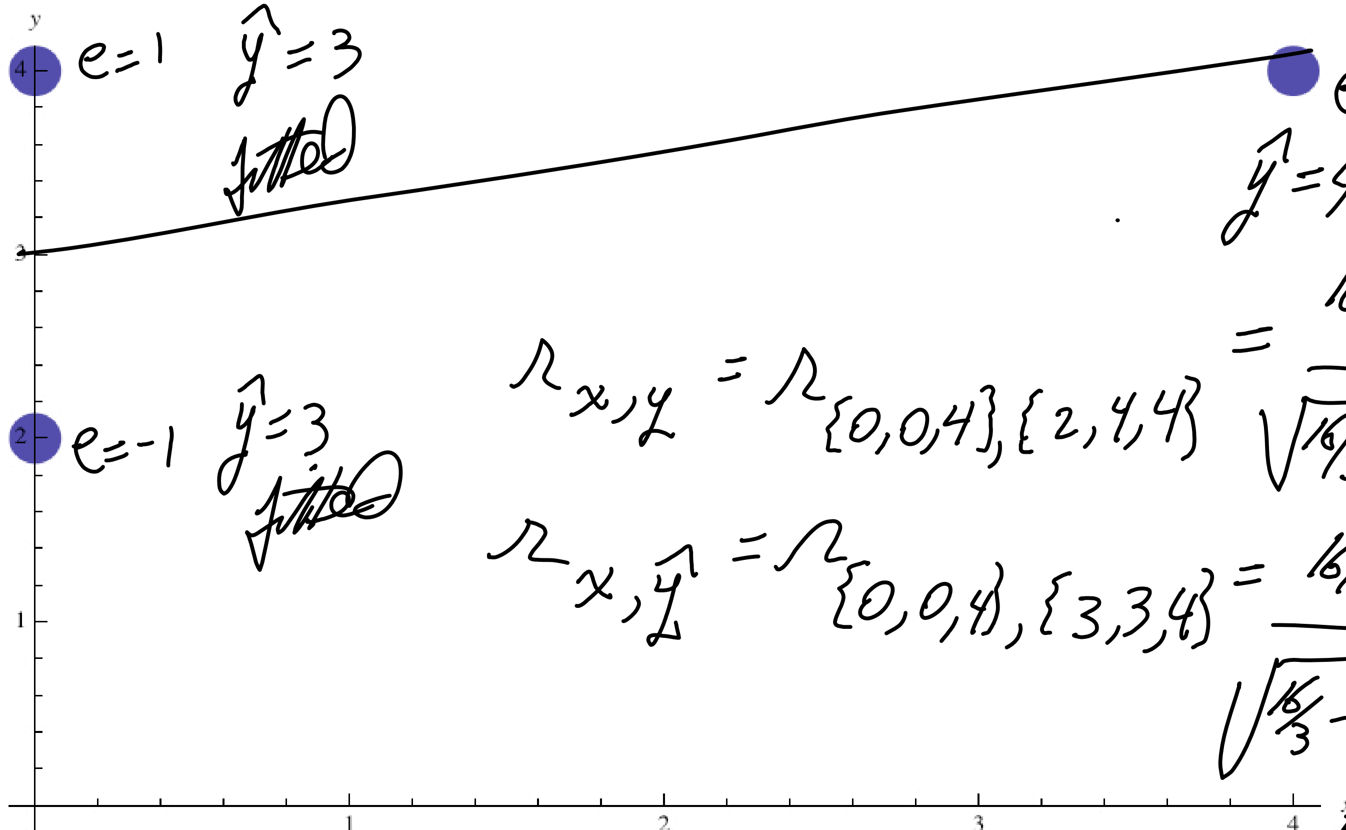
28. Verify that $r^2 = 1 - \frac{s_e^2}{s_y^2}$ $\leftarrow 1$, $\leftarrow 1.1547^2$

r^2 works out to .25

$r = .5$

$$s_e^2 = s^2_{\{-1,1,0\}} = \frac{2}{2} = 1 \qquad s_y^2 = 1.1547^2$$

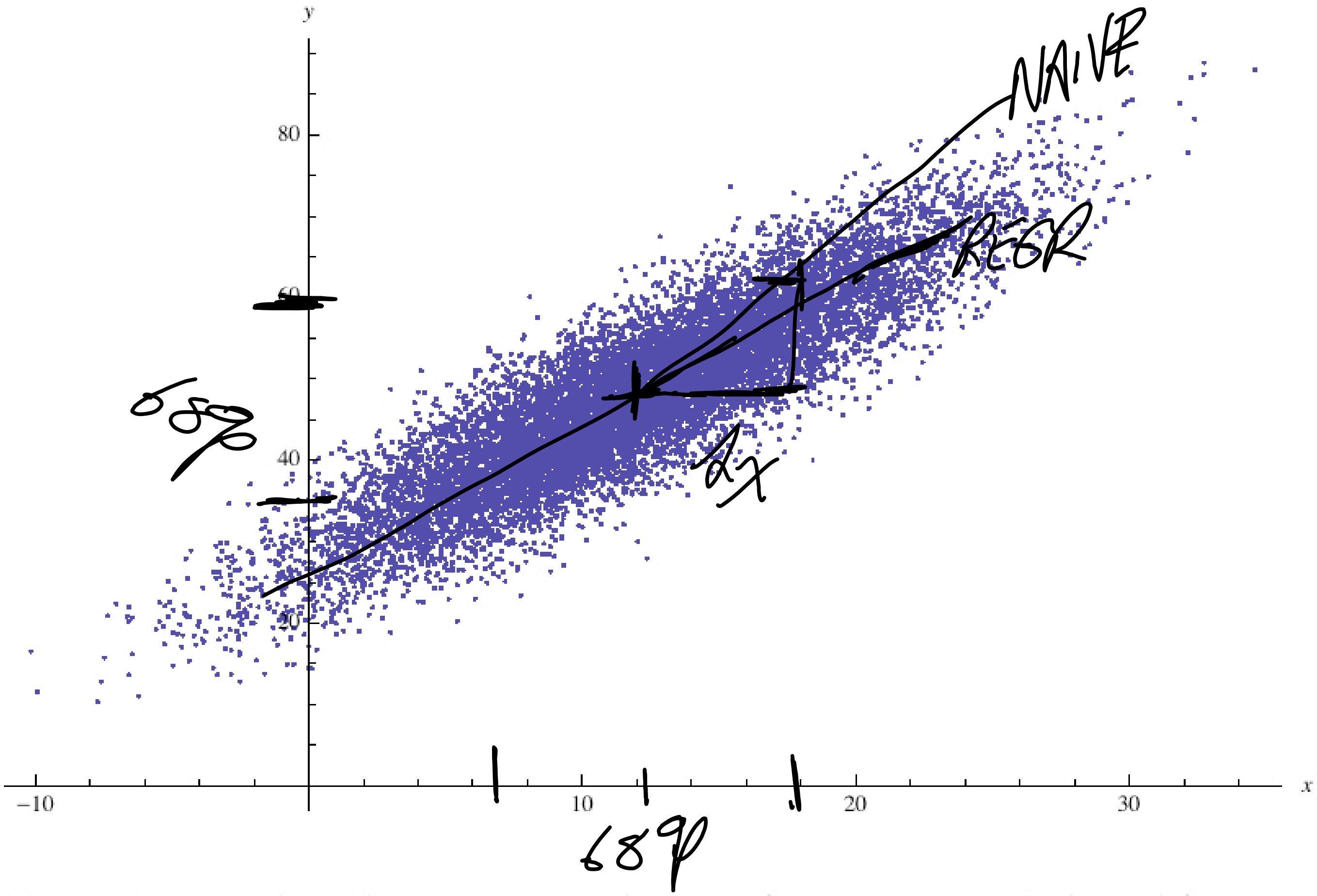
29. Interestingly, the correlation of x with the fitted values is exactly equal to |r|. Verify that it is so in this case.



$$r_{x,y} = r_{\{0,0,4\},\{2,4,4\}} = \frac{16/3 - 4/3\ \frac{10}{3}}{\sqrt{16/3-(4/3)^2}\sqrt{\frac{36}{3}-\left(\frac{10}{3}\right)^2}}$$

$$r_{x,\hat{y}} = r_{\{0,0,4\},\{3,3,4\}} = \frac{16/3 - 4/3\ \frac{10}{3}}{\sqrt{\frac{16}{3}-(4/3)^2}\sqrt{34/3-\left(\frac{10}{3}\right)^2}}$$

EXACTLY THE SAME



30. Draw in the regression of y on x. Identify and label $\overline{x}$, $\overline{y}$, s_x, s_y (using 68% rule).

$-\qquad \bar{x} \qquad \bar{y} \qquad \overline{x^2} \qquad \overline{y^2} \qquad \overline{xy}$

34.01 456. 1194.58 208788. 15391.9

NEED n

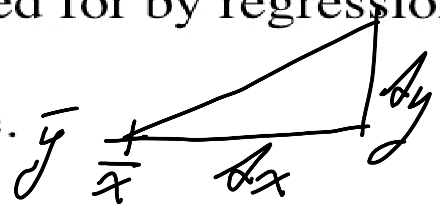
SUPPOSE n LARGE

1. $s_{y\cdot} = \cancel{\sqrt{\frac{n}{n-1}}}\underbrace{\sqrt{208788-456^2}}_{\hat{\sigma}_y}$ n LARGE

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

3. The fraction of s_y^2 accounted for by regression on x.

4. The slope of the naive line.



SLOPE $\dfrac{\Delta y}{\Delta x} = \dfrac{\hat{\sigma}_y}{\hat{\sigma}_x}$

5. The slope of the regression line of y on x. : $r\,\Delta y/\Delta x$ REGR LINE MORE HORIZONTAL

6. The slope of the regression line of x on y (which would apply if the variables were interchanged).

$$r = \frac{15392 - 34(456)}{\sqrt{1195-34^2}\sqrt{208788-456^2}}$$

ASIDE

$\overline{xy}-\bar{x}\bar{y}$

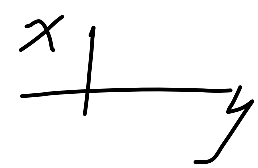
FOR $y = x$ GET

$\overline{x^2}-\bar{x}^2$

#3. ANS. $r^2 = 1 - \dfrac{\Delta^2_{RESID}}{\Delta_y^2}$ SHORT ANS. IF $r = .9$ YOU'VE EXPLAINED $.9^2 \sim 81\%$ OF Δ_y^2.

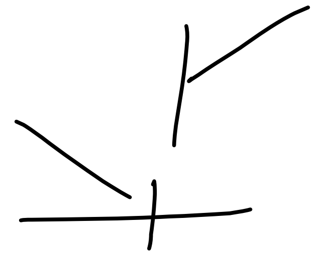
#6. REGR y ON x (USUAL) REGR SLOPE IS $r\,\Delta y/\Delta x$

FLIP TO REGR x ON y " " " $r\,\Delta x/\Delta y$.



TIME $\bar{x}$ CAL $\bar{y}$ $\overline{x2}$ $\overline{y2}$ $\overline{xy}$

34.01 456. 1194.58 208788. 15391.9

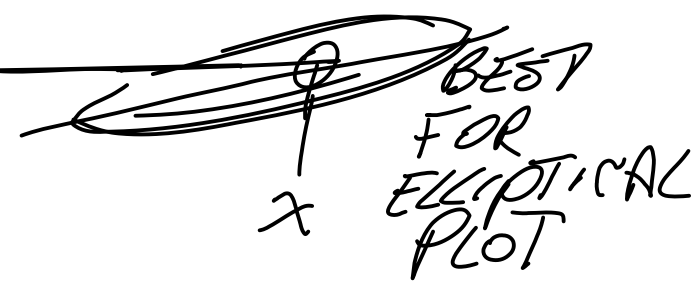


7. r[2 x - 4, 6 y + 2] 2,6 HAVE SAME SIGN

8. r[-x + 2, y - 6] $= -r[x, y]$ −1, 1 OPPOSITE SIGN

9. For an ELLIPTICAL plot having the above averages, the aver age calories for all subjects having time 36. $\approx \bar{x}$ SO AVG y FOR SUCH $x = 36 \approx \bar{y} = 456$.

10. For an ELLIPTICAL plot having the above averages, the best (by least squares) prediction for calories for a student with time 36. IS PT ON REGR LINE SO ≈ 456



11. The independent variable.
ALWAYS x (= TIME)

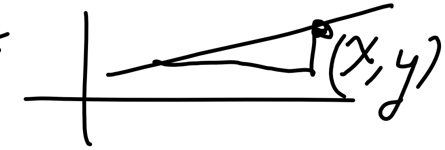
12. The dependent variable.

y —

13-21. r[x, y] = 0.9, $s_x = 2$, $s_y = 5$, $\overline{x} = 22$, $\overline{y} = 54$.

13. Determine r[y, x]. $= r[x, y] = 0.9$ SYMMETRIC

14. For points (x, y) on the regression line determine the numerical value of $\dfrac{y-\overline{y}}{x-\overline{x}}$ $= r \dfrac{s_y}{s_x} = 0.9\left(\dfrac{5}{2}\right)$ SLOPE OF REGR.



15. For x = $\overline{x} + s_x$ the regression prediction of y is $\overline{y} + (?)\, s_y$.

$r = 0.9$

16. For x = 18 the regression prediction of y is?

17. Regression predictions (15), (16) are sometimes useful even if the plot is not elliptical. If the plot IS ELLIPTICAL what is the special nature of the plot of vertical strip averages?

#16. PRED y FOR $x=18$ $\quad \dfrac{y-\overline{y}}{18-\overline{x}} = .9\left(\dfrac{5}{2}\right)$ SLOPE

$\text{PRED}(y) = \underset{\overline{y}}{54} + (18-\underset{\overline{x}}{22})\,.9\,(5/2).$

#17. PLOT OF VERTICAL STRIP AVGS.!

$r[x, y] = 0.9$, $s_x = 2$, $s_y = 5$, $\overline{x} = 22$, $\overline{y} = 54$.

18. If the plot is ELLIPTICAL what is the average y-score for all (x, y) pairs with x = 18?

19. If the plot is ELLIPTICAL what is the standard deviation of y-scores for all (x, y) pairs with x = 18?

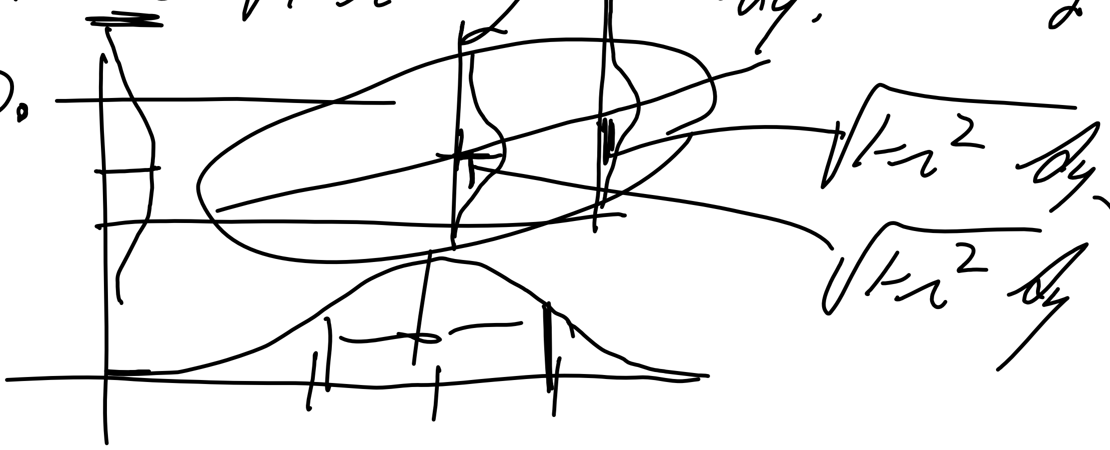
20. If the plot is elliptical, sketch the distribution of x, and the distribution of y.

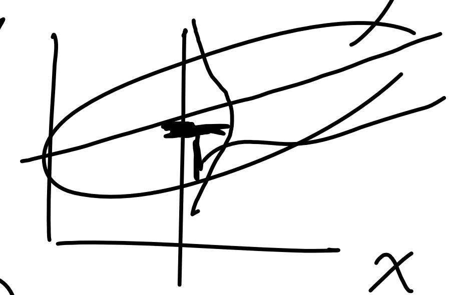
21. Draw a picture illustrating all of (18), (19), (20).

#18. POINT ON REGR AT $x = 18$ (is $\bar{y} + (18 - \bar{x})\, r \, {}^{\sigma_y}/_{\sigma_x}$)

#19. ANS. $\sqrt{1-r^2}\,\sigma_y \ll \sigma_y$.

#20.





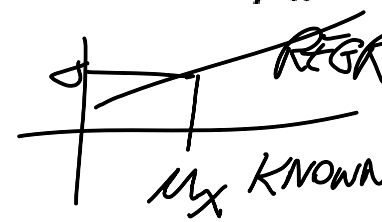
22-23. r[x, y] = 0.9, $s_x = 2$, $s_y = 5$, $\overline{x} = 22$, $\overline{y} = 54$.

22. Using (14), if I tell you that the population mean of x is μ_x = 26 what is the regression-based estimate for μ_x?

ABOUT ESTIMATING μ_y WHEN μ_x IS KNOWN

$\bar{y} + (\mu_x - \bar{x})\, r\, s_y / s_x$

DOES NOT REQUIRE PLOT BE ELLIPTICAL



23. Give the 95% CI for the estimate (19) if n is large. (The plot need not be elliptical since the estimator (19) is dependent upon $\overline{x}$ and $\overline{y}$ which are approximately jointly normal distributed for large n.)

FOR ABOVE GET 95% CI

$$\underbrace{\bar{y} + (\underset{\text{KNOWN}}{\mu_x} - \bar{x})\, r\, s_y / s_x}_{\text{NEW ESTIMATOR}} \pm 1.96 \frac{s_y}{\sqrt{n}}\, \cancel{FPC} \underbrace{\sqrt{1-r^2}}$$

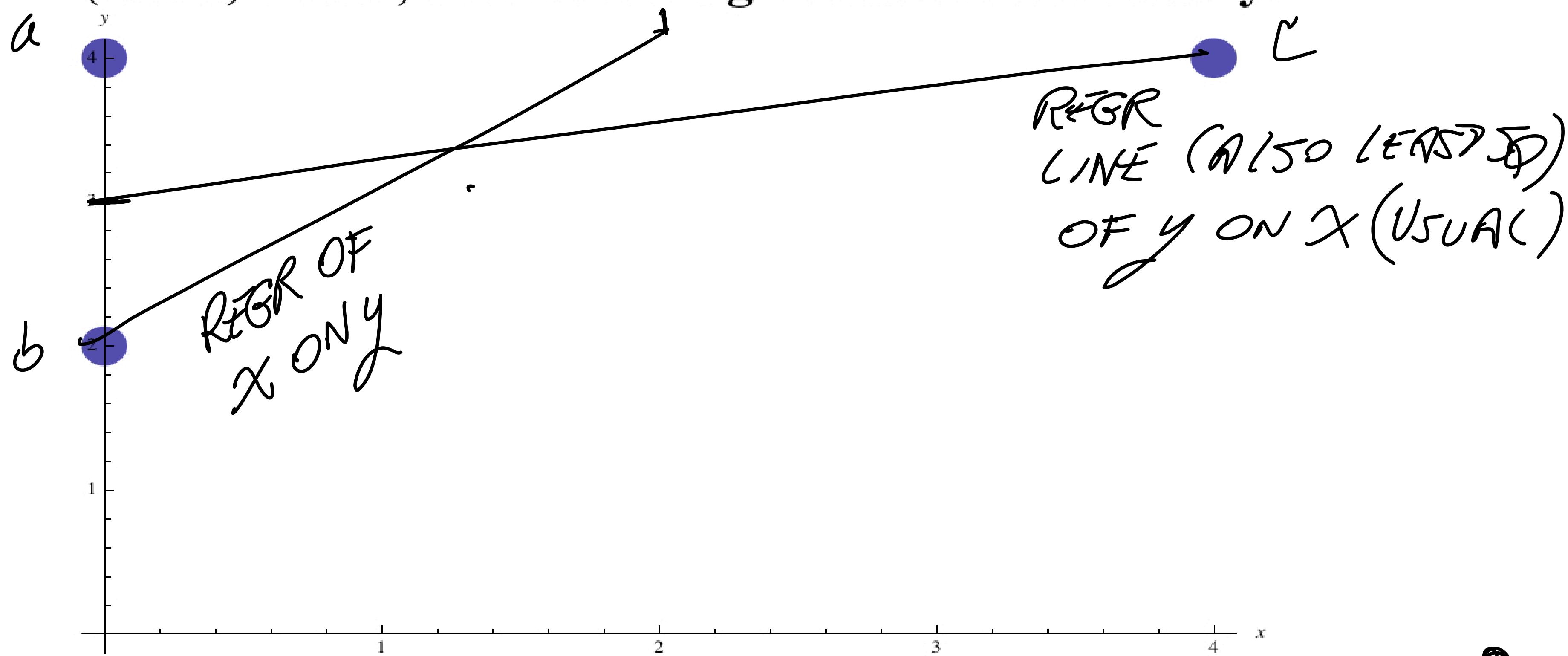
$n \to 4n$ ACHIEVES SAME

e.g. IF = $\frac{1}{2}$

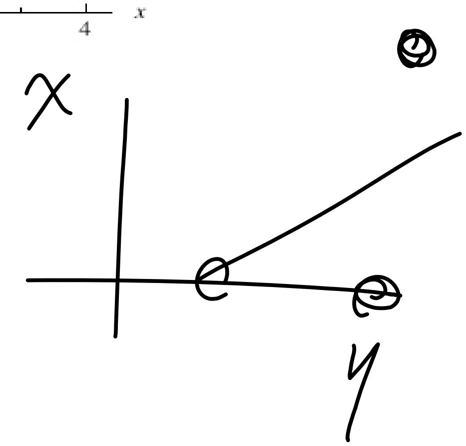
VS JUST USING y SCORES.

$$\bar{y} \pm 1.96\, s_y / \sqrt{n}\, \cancel{FPC}$$

24. For the plot below, sketch the regression line for y on x (usual). Also, sketch the regression line for x on y.



vs REGR OF x ON y.
DIFFERENT LINES



```
= regrtable[{0, 0, 4}, {2, 4, 4}]
```

atrixForm=

x	y	x^2	y^2	xy
0	2	0	4	0
0	4	0	16	0
4	4	16	16	16
—	—	—	—	—
1.33333	3.33333	5.33333	12.	5.33333

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}} = \frac{5.33 - 1.33 \; 3.33}{\sqrt{5.33 - 1.33^2}\sqrt{12 - 3.33^2}} = .5$$

25. For the plot just above calculate the slope of regression. Confirm it with what you see in the plot.

$$\text{SLOPE} = r\,\hat{\sigma}_y/\hat{\sigma}_x = \frac{5.33 - 1.33 \; 3.33}{5.33 - 1.33^2}$$

See that in general

26. From your calculations (25) what is s_y?

$$\sqrt{12 - 3.33^2}$$

27. Calculate s_e.

e ARE THE RESIDUALS OF REGRESSION

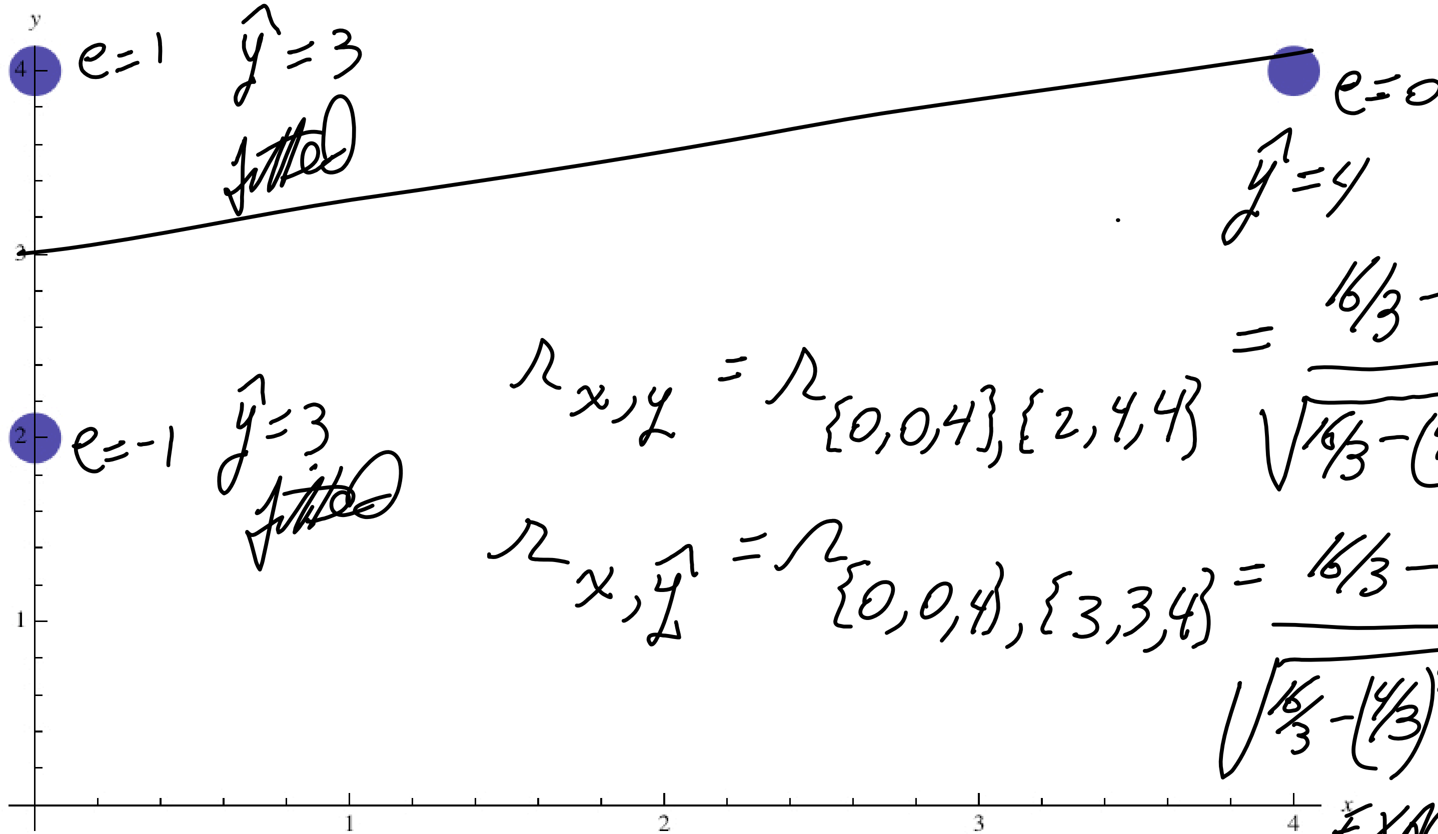
28. Verify that $r^2 = 1 - \frac{s_e^2}{s_y^2}$ $\leftarrow 1$, $\leftarrow 1.1547^2$

r^2 works out to .25

$r = .5$

$$s_e^2 = s^2_{\{-1,1,0\}} = \frac{2}{2} = 1 \qquad s_y^2 = 1.1547^2$$

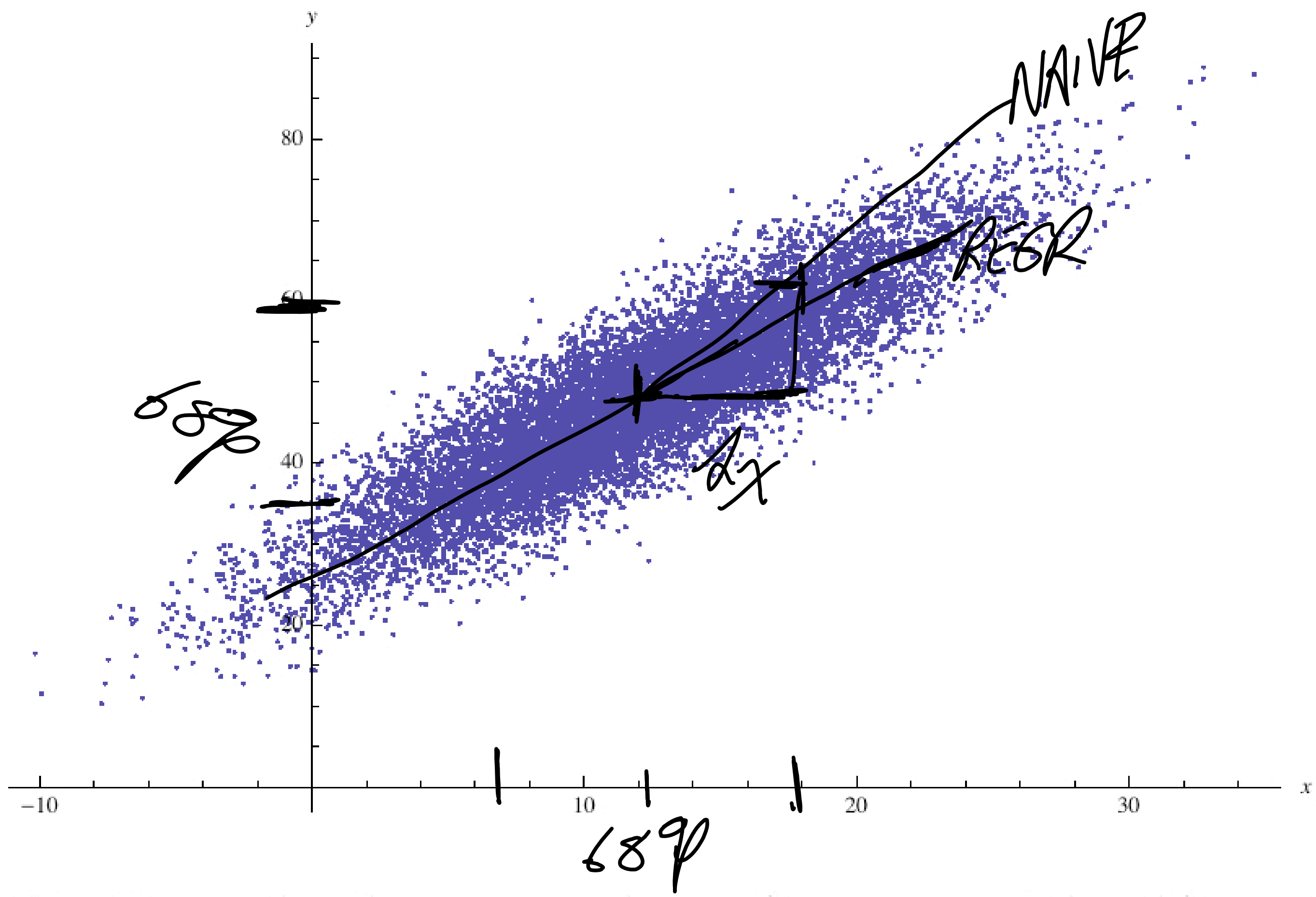
29. Interestingly, the correlation of x with the fitted values is exactly equal to |r|. Verify that it is so in this case.



$$r_{x,y} = r_{\{0,0,4\},\{2,4,4\}} = \frac{16/3 - 4/3 \cdot \frac{10}{3}}{\sqrt{16/3 - (4/3)^2}\sqrt{\frac{36}{3} - \left(\frac{10}{3}\right)^2}}$$

$$r_{x,\hat{y}} = r_{\{0,0,4\},\{3,3,4\}} = \frac{16/3 - 4/3 \cdot \frac{10}{3}}{\sqrt{\frac{16}{3} - (4/3)^2}\sqrt{34/3 - \left(\frac{10}{3}\right)^2}}$$

EXACTLY THE SAME



30. Draw in the regression of y on x. Identify and label $\overline{x}$, $\overline{y}$, s_x, s_y (using 68% rule).