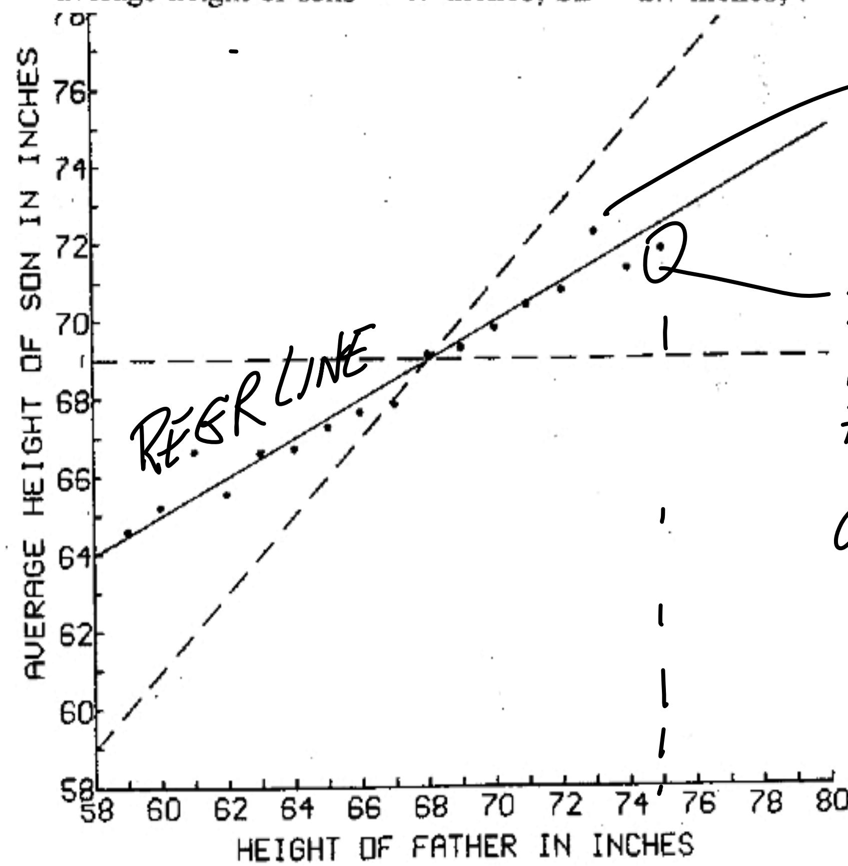
ASIDE:
REGR LINE
LEAST SQUARES LINE

average height of fathers ≈ 68 inches, SD ≈ 2.7 inches
average height of sons ≈ 69 inches, SD ≈ 2.7 inches, $r \approx 0.5$



ALL VERTICAL STRIP AVES.

$x = 75$ POSSIBLE FATHER HEIGHT
y = AVE HEIGHT OF SONS WHOSE FATHERS HTS WERE ≈ 75

" Freedman Pisani Purves, 1980 "

— STATISTICS —

REGR LINE ≈ PLOT OF VERT STRIP AVES IS PARTICULAR TO ELLIPTICAL PLOTS.

$\bar{x} = 19.56$

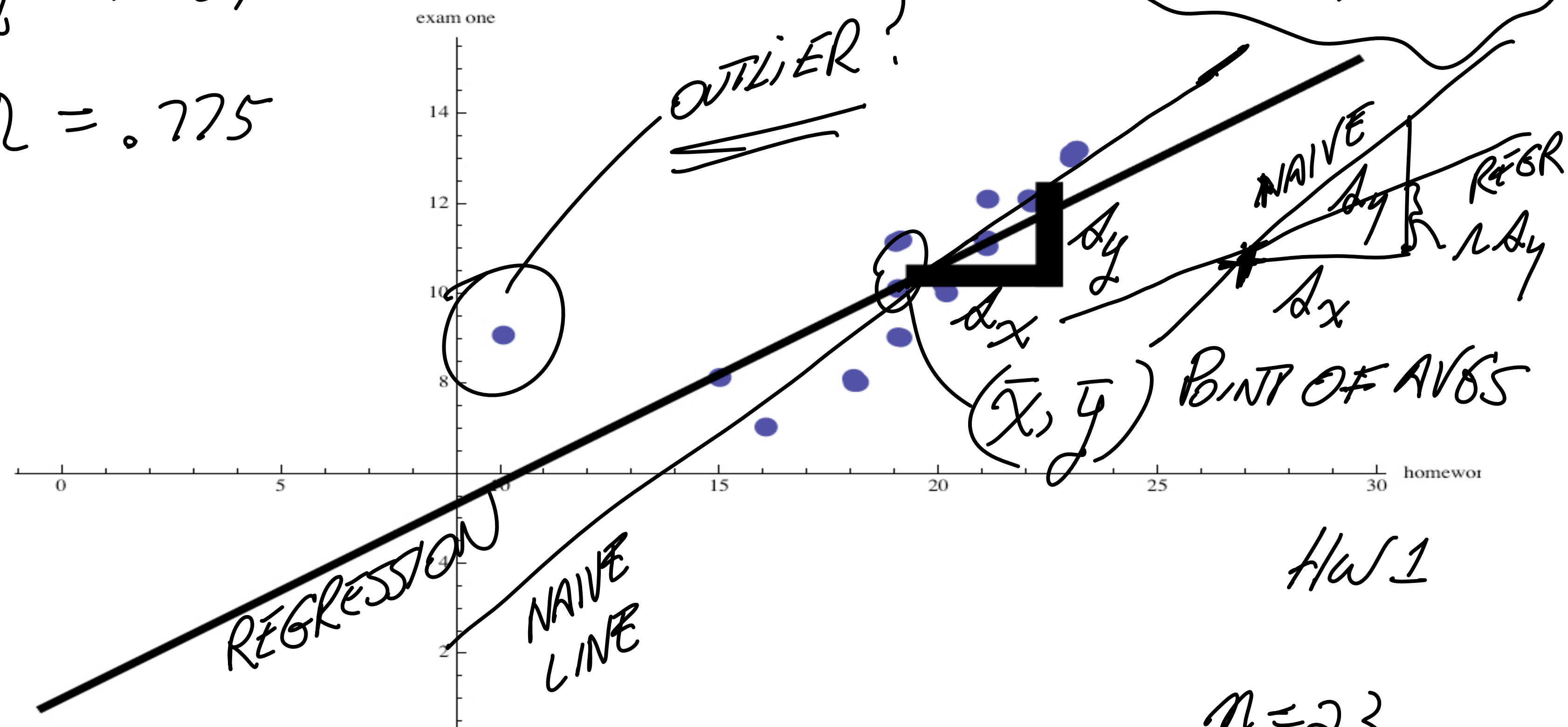
$\bar{y} = 10.39$

$r = .775$

EXAM 1

$r[x, y] = r[ax+b, cy+d]$

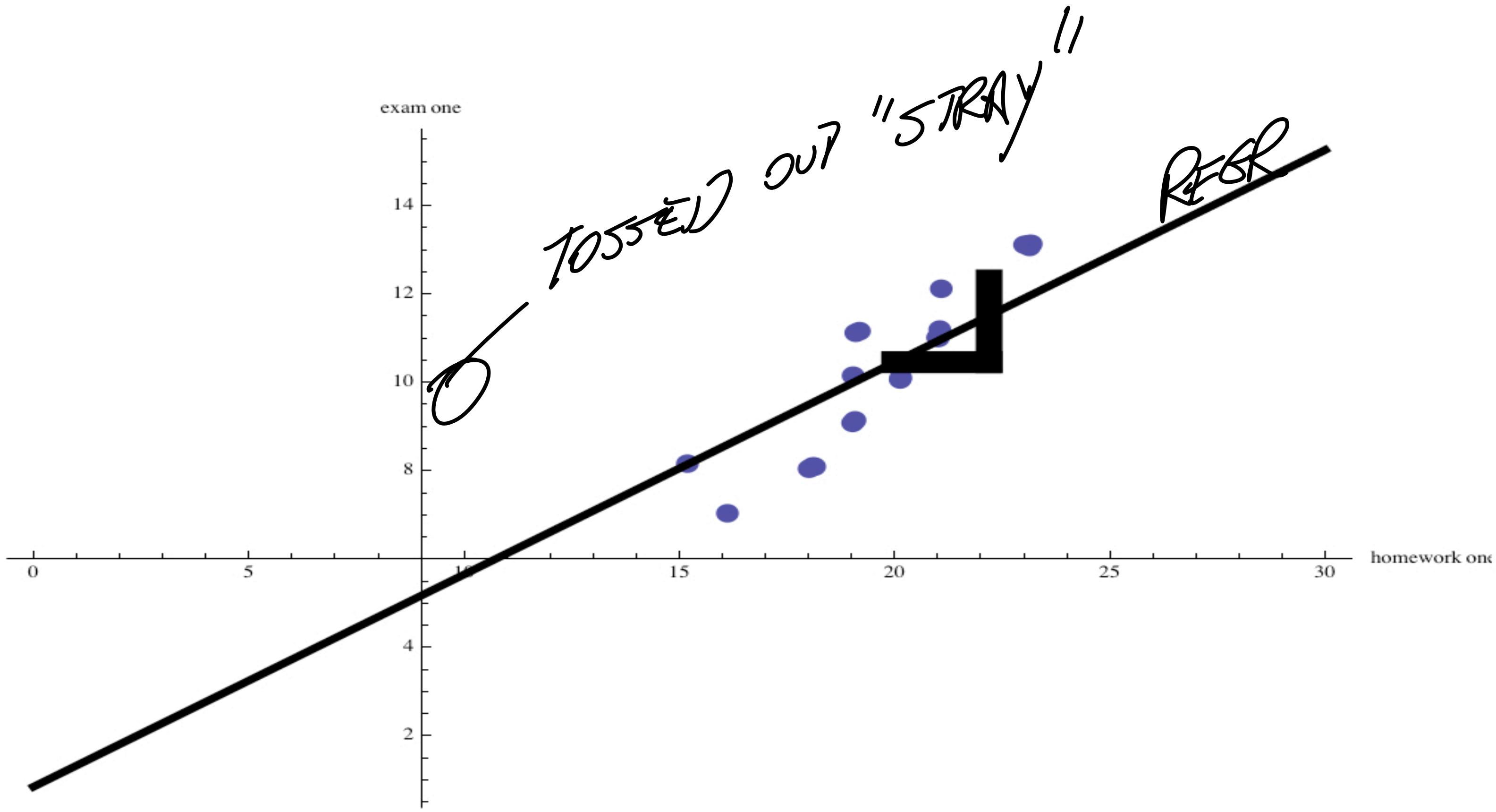
$a>0, b>0$



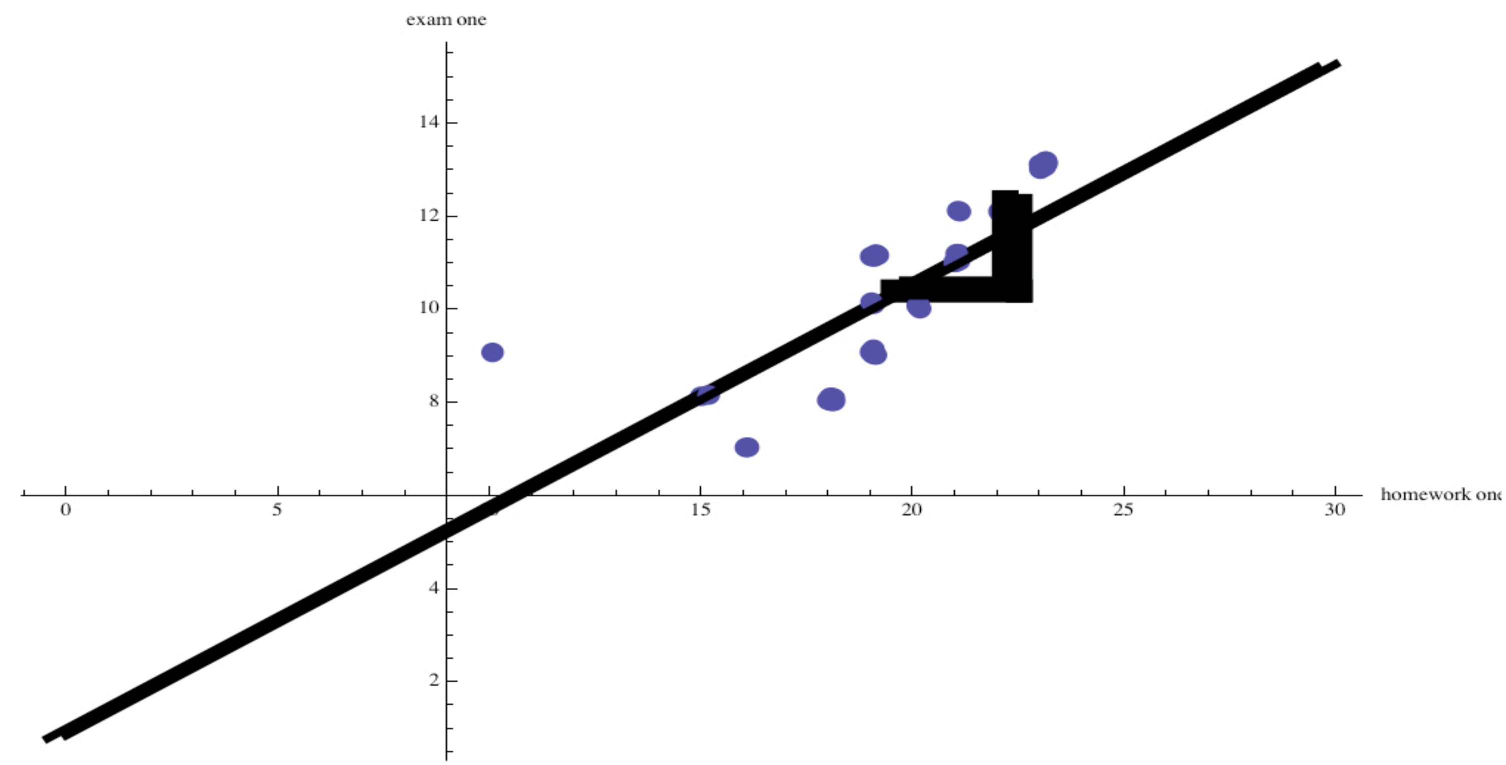
HW 1

$n = 23$
RANDOMLY
SELECTED
STUDENTS

{Mean[hw1], Mean[e1]} 1.0	**{19.5652, 10.3913}**
{σ[hw1], σ[e1]} 1.0	**{2.96099, 1.83538}**
r[hw1, e1] 1.0	**0.775338**
r[hw1, e1] σ[e1]/σ[hw1] 1.0	**0.480595**



{Mean[hw1a], Mean[e1a]} 1.0 **{19.5652, 10.3913 }**
{σ[hw1a], σ[e1a]} 1.0 **{2.96099, 1.83538}**
r[hw1a, e1a] 1.0 **0.775338**
r[hw1a, e1a] σ[e1a]/σ[hw1a] 1.0 **0.480595**



Lines of regression with and without the point (10, 9).

FOR ELLIPTICAL PLOT -

IF FATHER'S HT IS σ_x ABOVE $\bar{x}$: $(\bar{x} + \sigma_x)$

& IF r (CORRELATION) IS $r = .8$

THEN, THE <u>AVG HT</u> OF SONS BORN

TO FATHERS OF HT $\bar{x} + \sigma_x$ IS $r \, 1 SD_y + \bar{y}$

ie $\bar{y} + .8 \sigma_y$

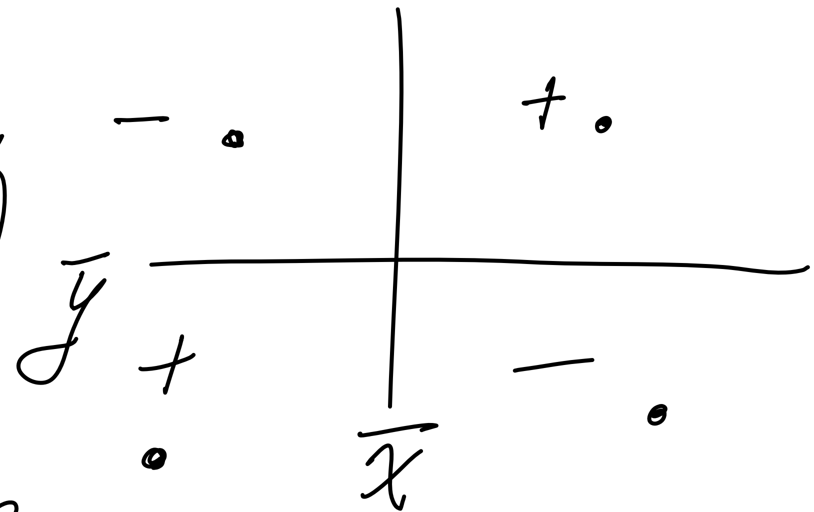
ANOTHER - IF WE LOOK AT ALL STUDENTS WHOSE EXAM 1 SCORE IS $\bar{x} + 2\sigma_x$, THEN THEIR AVG SCORE ON EXAM 2 IS $\approx \bar{y} + r \, 2 \sigma_y$

$= \bar{y} + 1.6 \sigma_y$

WHERE CORRELATION CAME FROM.

IDENTIFY PLOT THAT IS
LIKE A LINE

$$\boxed{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\Big/ n-1}{s_x\, s_y}} = r$$



INSENSITIVE TO LOC + (POS) SCALE CHANGES.

$|r| \le 1$ i.e. $-1 \le r \le 1$

~~Fix!~~

$|r| = 1 \iff$ ALL THE POINTS LIE PERFECTLY ON (THE REGR) LINE

COINCIDENT W/ NAIVE LINE IN THIS $r = \pm 1$ CASE

NOTE: $\left(\overline{xy} - \bar{x}\,\bar{y}\right) \Big/ \left(\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}\right)$ CAREFUL!

STT200 CH 7 (CONT)

EXAM 1 GRADE $= 2 + 0.5(\text{SCORE} - 8)$

eg SCORE 8 → GRADE $= 2 + .5(8-8) = 2$

10 → $= 2 + .5(10-8) = 3$

12 → 4

>12 Keep 4.

WILL SEND & POST HW DUE NEXT TU.

SON y

GALTON'S DATA -

$$s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$

x = FATHER HT

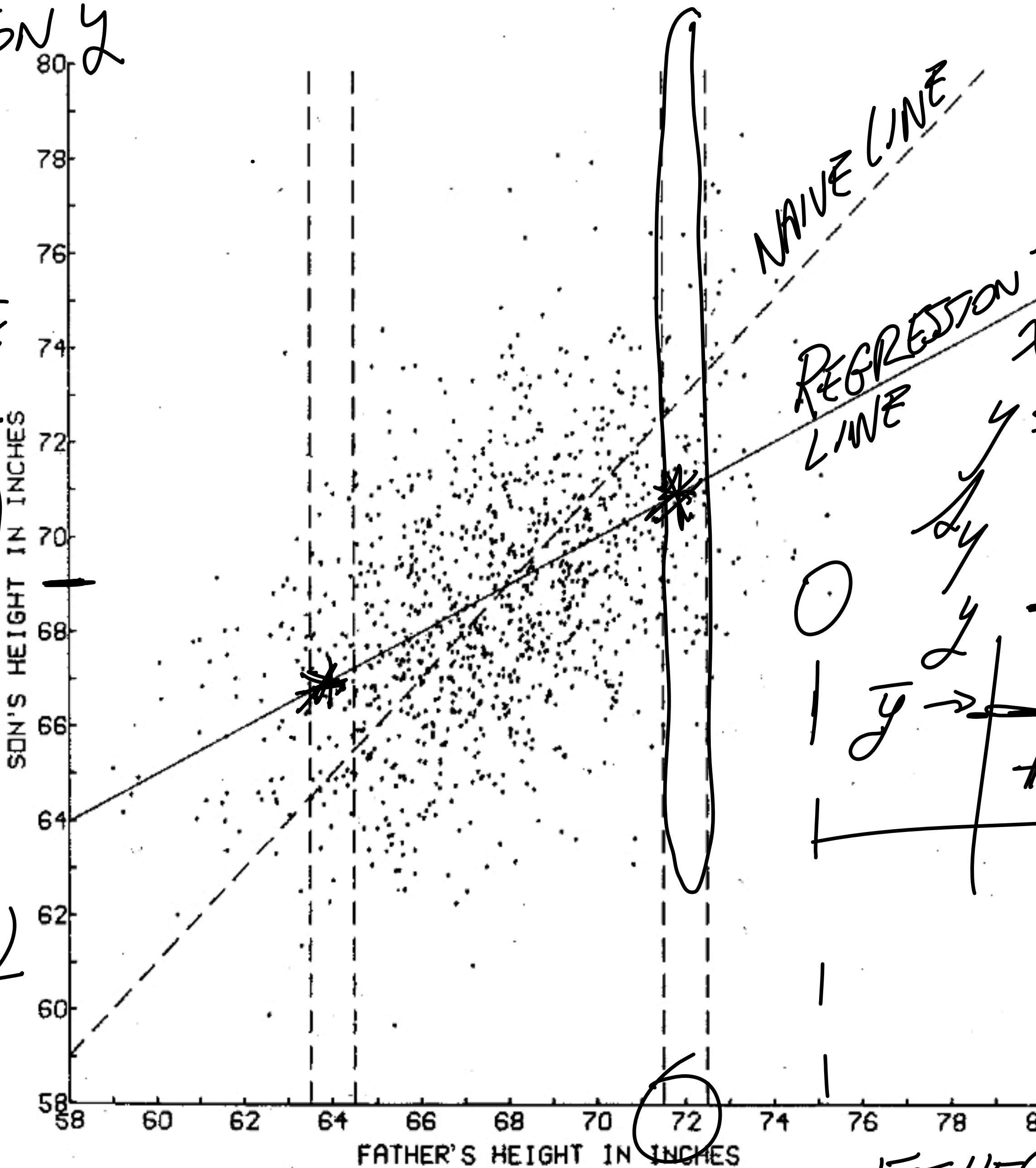
y = SON HT

1st x 1st $\bar{x}$

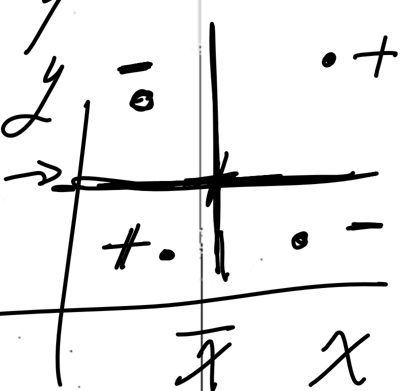
$$\frac{\frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}}{s_x \; s_y}$$

~2x ~2x

$$\frac{(x-\bar{x})(\quad)}{s_x}$$



s_y



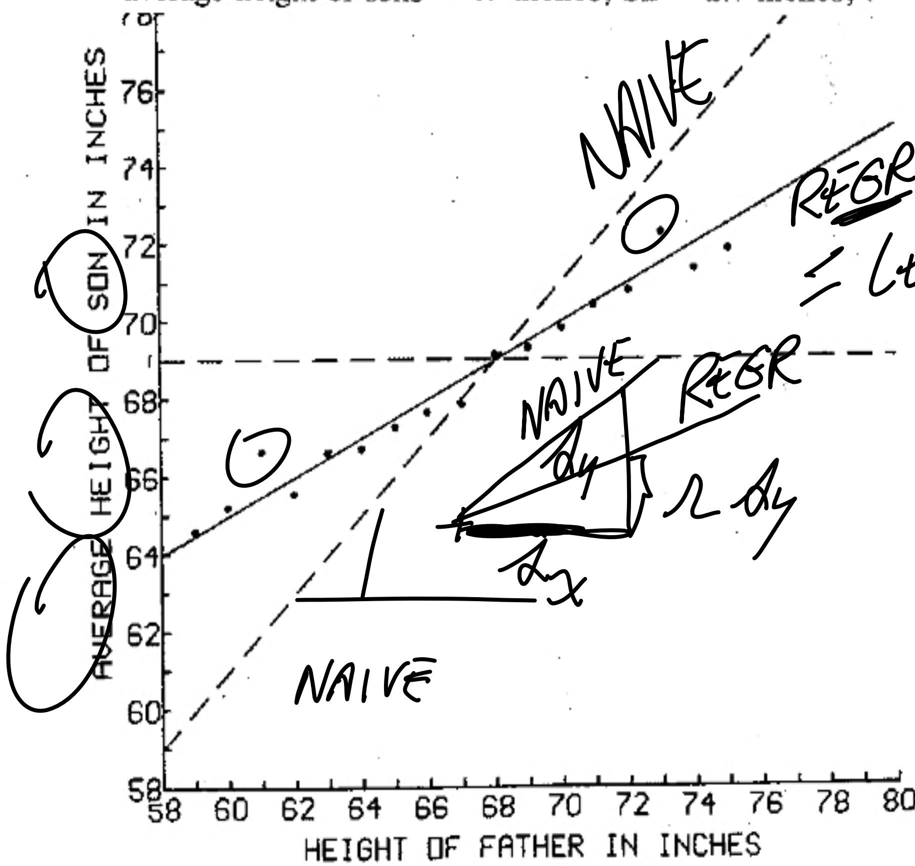
FATHER X

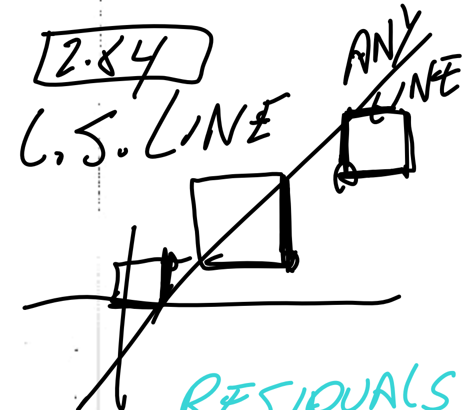
INFLUENCES INCL DIET, GENES, \$.

$$r = \frac{\left[\frac{\sum_i (x-\bar{x})(y-\bar{y})}{n-1}\right]}{s_x s_y}$$

$$r[5x+2,\ 3y-6] = r[x,y]$$

average height of fathers ≈ 68 inches, SD ≈ 2.7 inches
average height of sons ≈ 69 inches, SD ≈ 2.7 inches, $r \approx 0.5$



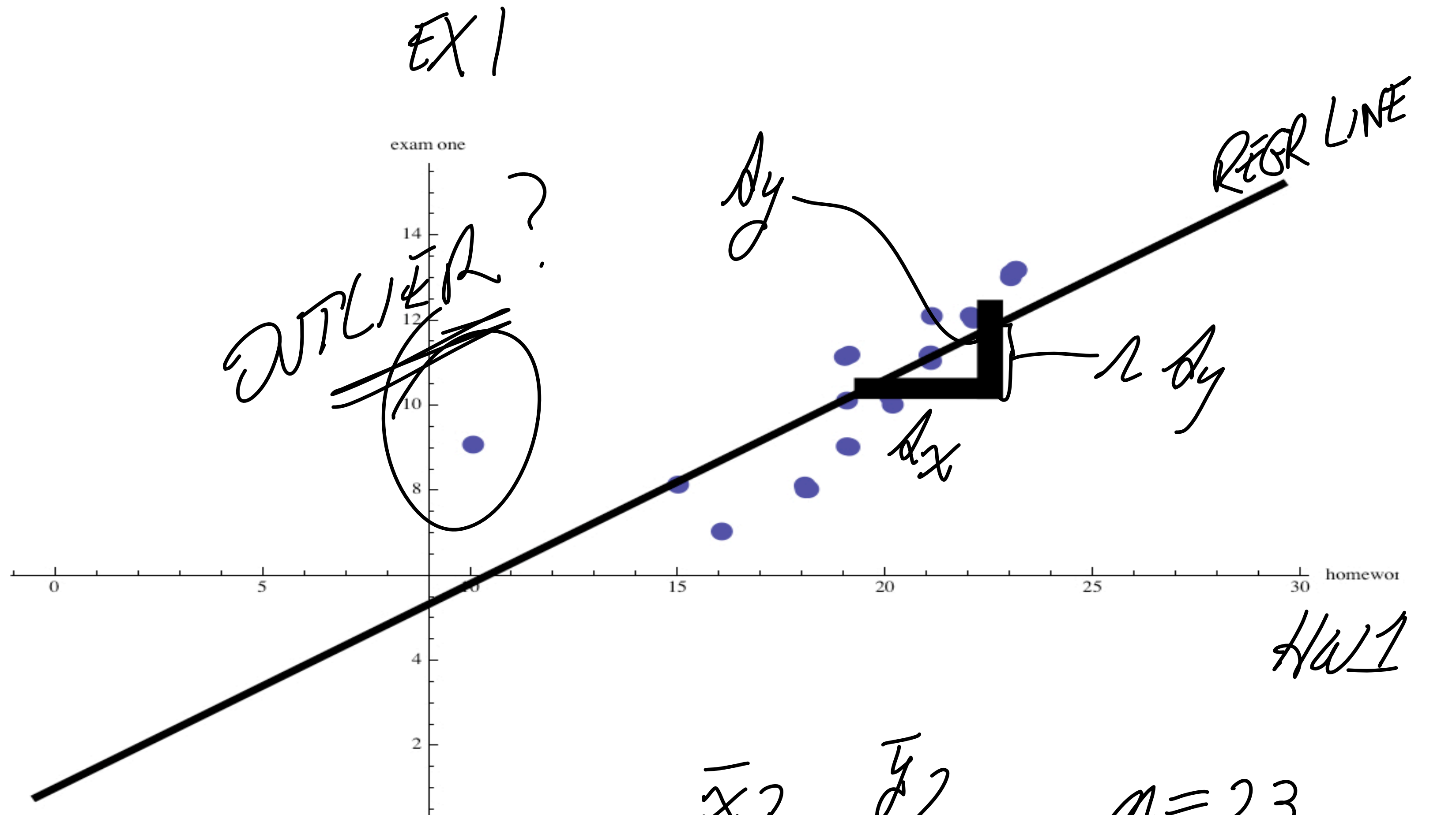


RESIDUALS (ON VERTICAL)
SUM SQUARES OF VERTICAL DISCREPANCIES

Freedman Pisani Purves, 1980

CLAIM: FOR ELLIPTICAL PLOTS THE VERTICAL STRIP AVGS PLOT ≈ AS REGRESSION LINE.

FOR EVERY PLOT LS LINE = REGR LINE



$\bar{x}$ ≈ $\quad\bar{y}$ ≈

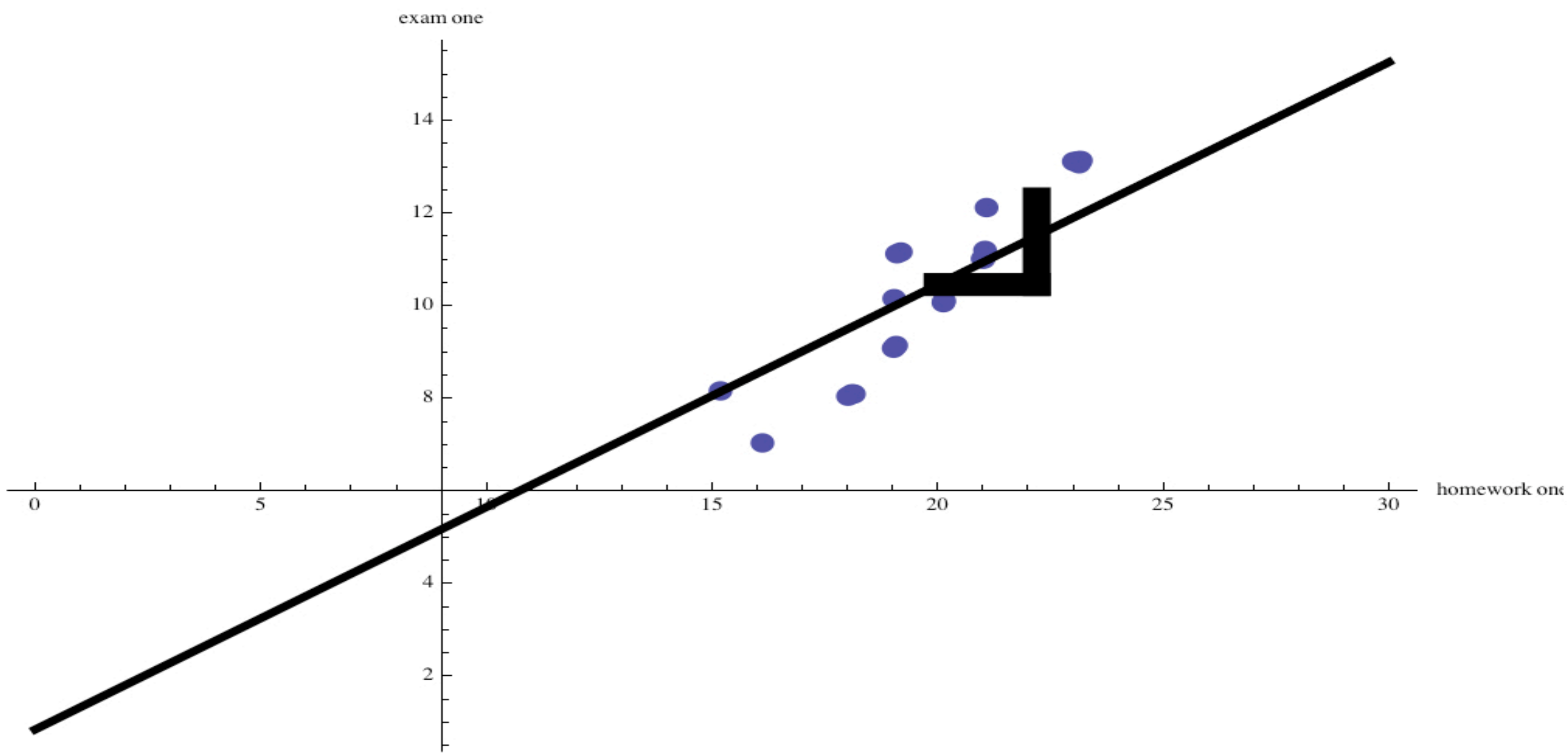
{Mean[hw1], Mean[e1]} 1.0	{19.5652, 10.3913}
{σ[hw1], σ[e1]} 1.0	{2.96099, 1.83538}
r[hw1, e1] 1.0	0.775338
r[hw1, e1] σ[e1]/σ[hw1] 1.0	0.480595

$n = 23$

$\sigma_x = 2.96 \quad \sigma_y = 1.83$

$r = 0.775$

IF YOU WERE $2\,\sigma_x$ ABOVE $\bar{x}$ (ON HW1) REGR LINE PREDICTS YOU WILL BE $r(2\sigma_y) = 2(0.775)\sigma_y$ ABOVE $\bar{y}$.

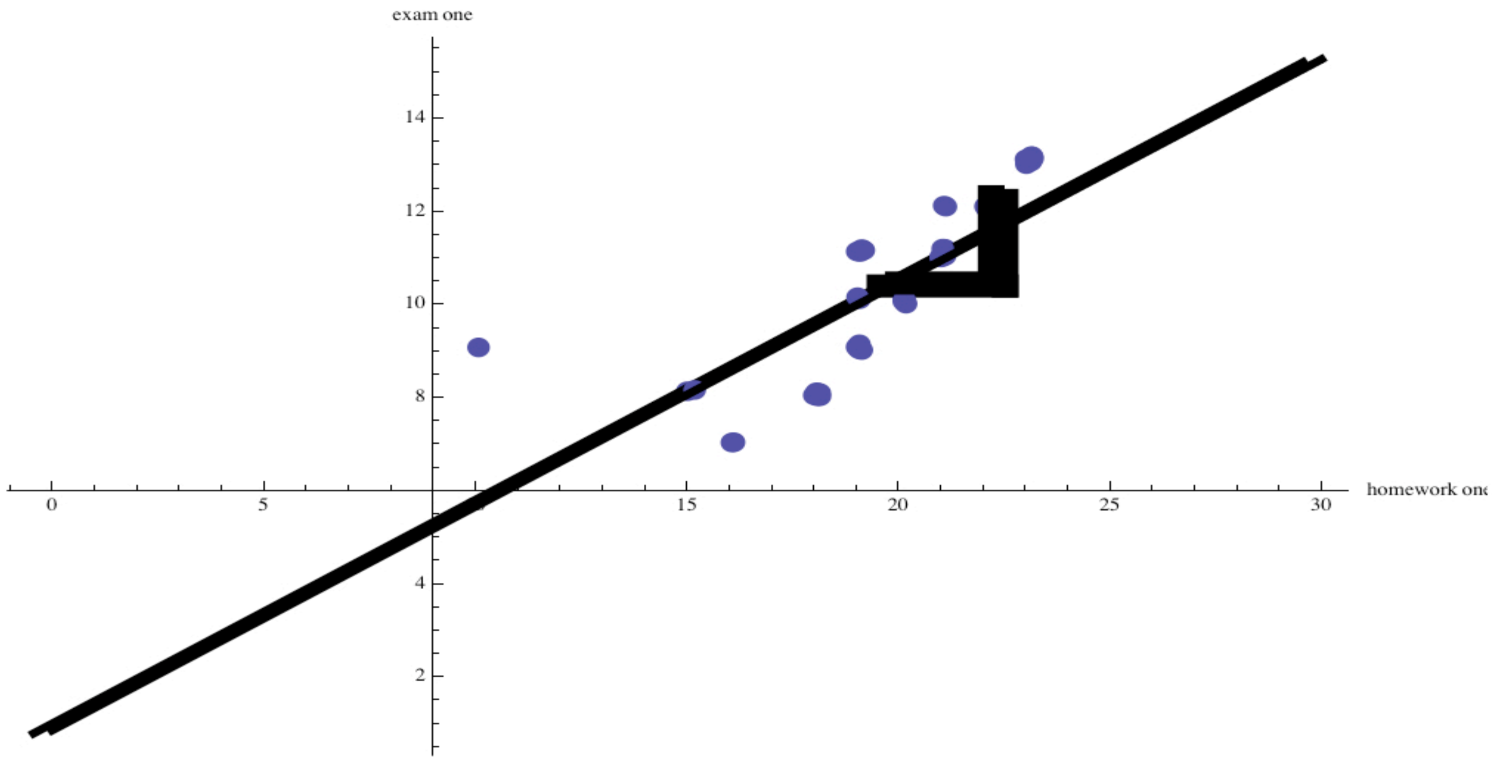


{Mean[hw1a], Mean[e1a]} 1.0 {19.5652, 10.3913 }

{σ[hw1a], σ[e1a]} 1.0 {2.96099, 1.83538}

r[hw1a, e1a] 1.0 0.775338

r[hw1a, e1a] σ[e1a]/σ[hw1a] 1.0 0.480595



Lines of regression with and without the point (10, 9).

DEF $\bar{r}$ & CALC OF r.

$$r = \frac{\sum_i (x-\bar{x})(y-\bar{y}) / (n-1)}{s_x \, s_y}$$

GOOD TO ROUNDING ERRORS —

CAUTION

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

(AVG OF PRODUCTS) — (PROD OF AVES)

$\overline{x^2}$ = avg of squares

$\bar{x}^2$ = sq of avg,

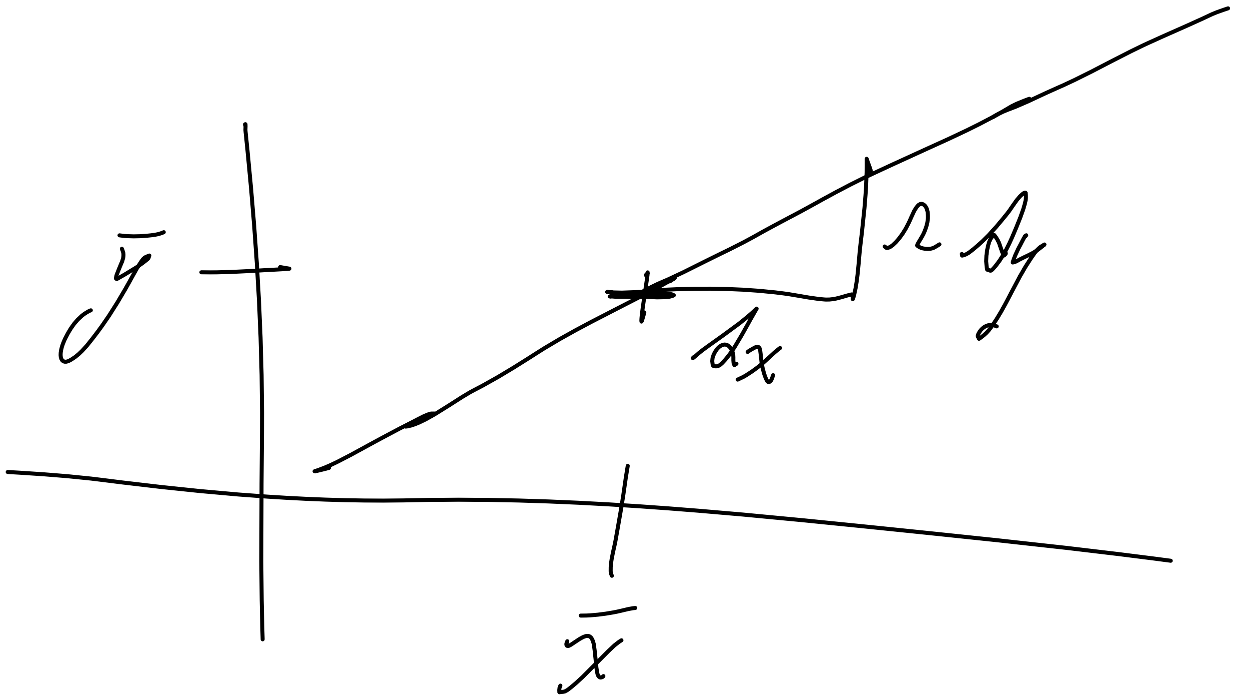
$$\boxed{s_y = \sqrt{3/2}\sqrt{15 - 3^2} = \sqrt{\tfrac{3}{2}}\sqrt{6} = 3}$$

	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

$n = 3$

$$s_x = \sqrt{\frac{3}{3-1}}\sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{3/2}\sqrt{3 - 1^2} = \sqrt{3}$$

$$r = \frac{0 - 1(3)}{\sqrt{3} \quad 3}$$



PLOT (x,y) PAIRS –