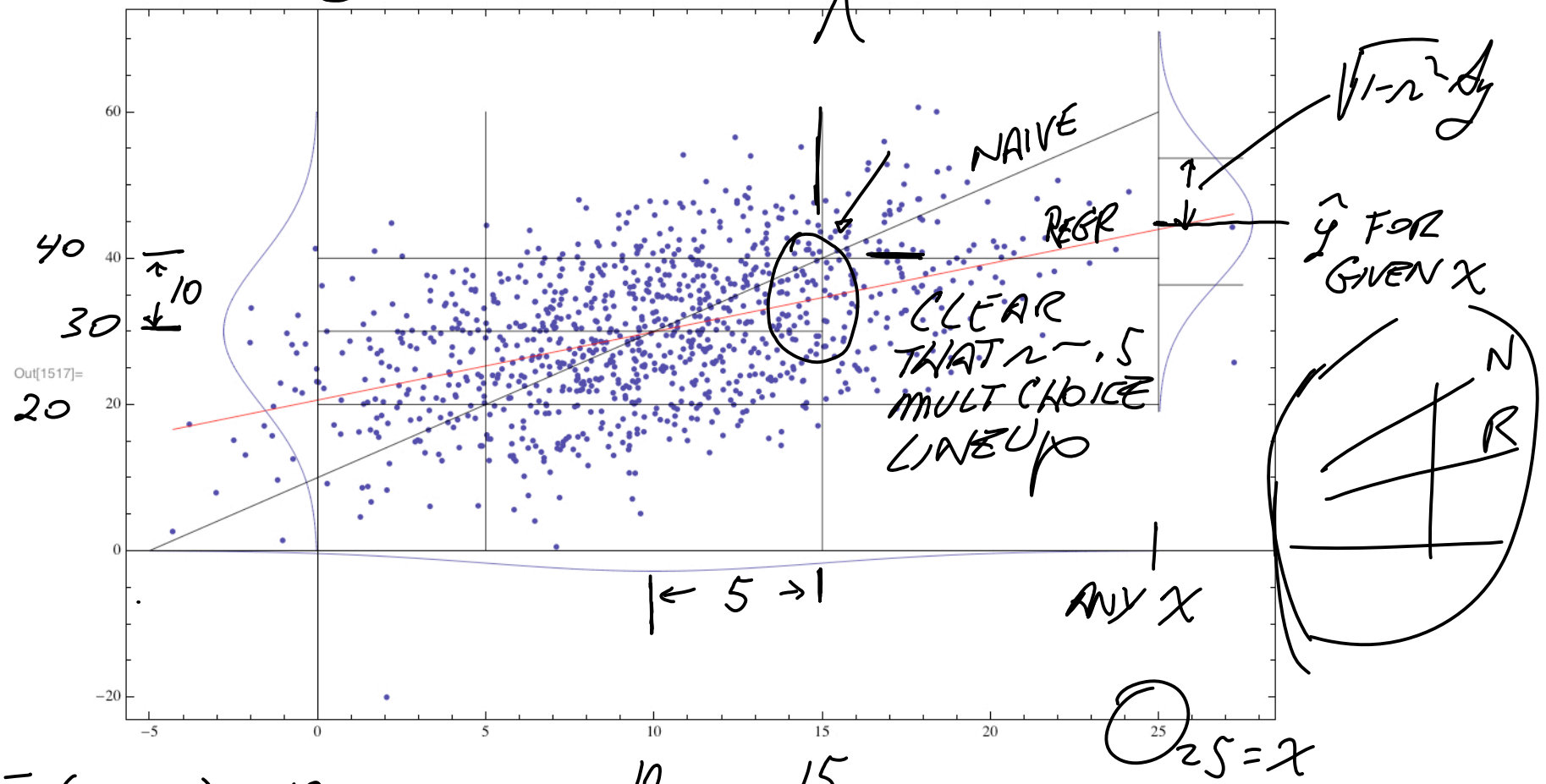
Stat 200 3-2-09

In[1517]:= regrpic[10, 30, 5, 10, 0.5, 1000, 70, .005]

(.5) n SOMESCALE DOT SIZE

10 30 5 10 .5 1000 70 .005



$\sqrt{1-r^2}\,\Delta_y$

$\hat{y}$ FOR GIVEN X

$\bar{x}\ (\sim \mu_x) \approx 10$

$\Delta_x\ (\sim \sigma_x) \approx 5$

$\bar{y}\ (\sim \mu_y) \approx 30$

$\Delta_y\ (\sim \sigma_y) \approx 10$

10 15

MULT CHOICE → $r \approx .5$

SD OF y FOR $x = 25$ IS $\sqrt{1-(.5)^2}\,\Delta_y = \sqrt{.75}\,10$

$\left(\text{SLOPE}_{\text{NAIVE}}\right) = \Delta_y/\Delta_x = \frac{10}{5} = 2$ $\left(\text{SLOPE}_{\text{REGR}}\right) = r(2) = .5(2) = 1$

For Exam 2 calculate sigmaHAT values on questions 8, 9

Calc. $\hat{\sigma}_x = \sqrt{\bar{x^2} - \bar{x}^2} = \sqrt{27.75 - 4.25^2}$

$\hat{\sigma}_y = \sqrt{18 - 3^2} = 3$

```
regrtable[{1, 9, 2, 5}, {2, 2, 8, 0}]
```

MatrixForm=

x	y	x^2	y^2	xy
1	2	1	4	2
9	2	81	4	18
2	8	4	64	16
5	0	25	0	0
—	—	—	—	—
4.25	3.	27.75	18.	9.

STLINE SETUP

$$r = \frac{\bar{xy} - \bar{x}\bar{y}}{\sqrt{\bar{x^2} - \bar{x}^2}\sqrt{\bar{y^2} - \bar{y}^2}} = \frac{9 - (4.25)(3)}{\sqrt{27.75 - 4.25^2}\sqrt{(3)}}$$

RANGE OF r IS $[-1, 1]$.

MLR

$R = |r[x, y]| = r[y, \hat{y}]$ RANGE OF R IS $[0, 1]$

FRAC OF s_y^2 ACCOUNTED FOR BY REGR IS r^2 (= MLR R^2)

```
regrstats[{1, 9, 2, 5}, {2, 2, 8, 0}]
{4.25, 3., [redacted] -0.40161, -0.38709
(-0.40161)^2
0.161291
√(1 - .161291) 3.4641
3.17246
```

slope

r values

Or

$= r$

$s_y\sqrt{1-r^2}$

LR (SD LINE)

REGR LINE

$+ x$

MULTIPLE LINEAR REGRESSION MLR

SD OF y FOR FIXED $x_1, x_2, \ldots$

SD $\sqrt{1-R^2}\, s_y$

$x_1, x_2 \ldots$

MLR

(1) IF I REPORT $s_y = 11.6$ & $\hat{b}_0 = 27.3$ (CONST TERM)

$\hat{b}_1 = 3.1$ $\hat{b}_2 = -2.2$

? WHAT IS $\hat{y}$ (PREDICTION FOR y) WHEN $x_1 = 82$, $x_2 = 17$

SUBJECT HAVING THESE SCORES —

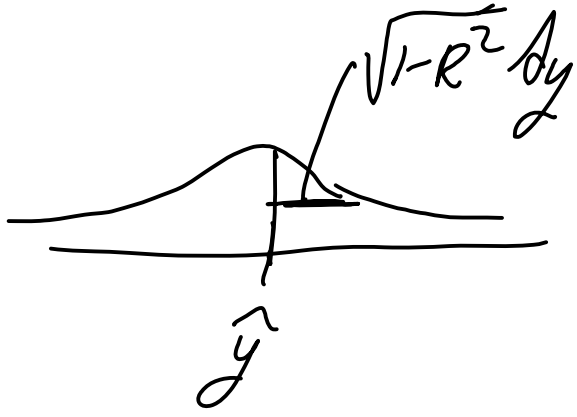
ANS. FEED IN THESE x-SCORES TO REGR COEFF.

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2$$

$$= 27.3 + 3.1(82) - 2.2(17)$$

[GEN'L REQUIREMENT IS $R \sim 1$
IF SO, MLR IS GOOD DESCRIPTOR]

[FOR BELL CURVE TO APPLY WE REQUIRE ELLIPTICAL PLOT.]



for given x_1, x_2

NEED MULT COR R.

QUES. $\hat{b}$? — YIELD — INPUTS H_2O, SOLAR, FERT etc.

MODEL $y_i = b_0 + b_1 x_{1i} + \cdots + b_k x_{ki} + \text{ERROR}$

LEAST SQUARES (REGRESSION)

FIND $\hat{b}_0, \ldots, \hat{b}_k$ (FROM DATA

THESE $\hat{b}_0, \ldots, \hat{b}_k$ ARE

$$\begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

CONST TERM

PRECISELY THOSE MINIMIZING

m_{yx}

m_{yy}

$$\sum_{i=1}^{1n} \left(y_i - \left(b_0 + b_1 x_{1i} + \cdots + b_k x_{ki} \right) \right)^2$$

WIGGLE THESE OF MINIMIZE THE SUMM

beta hat $[m_{yx}, m_{yy}]$

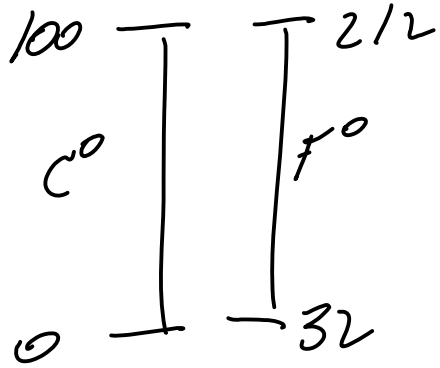
PERTAINING TO GENERALITIES

$\mathcal{L}[6x+3, 9y-4] = 2$

eg $\mathcal{L}[\text{LAX am temp}^{F^\circ}, \text{LAX pm temp}^{F^\circ}] = .6\ (\text{SAN})$

SAME FOR C° !!

$F^\circ = 32 + \frac{9}{5} C^\circ$

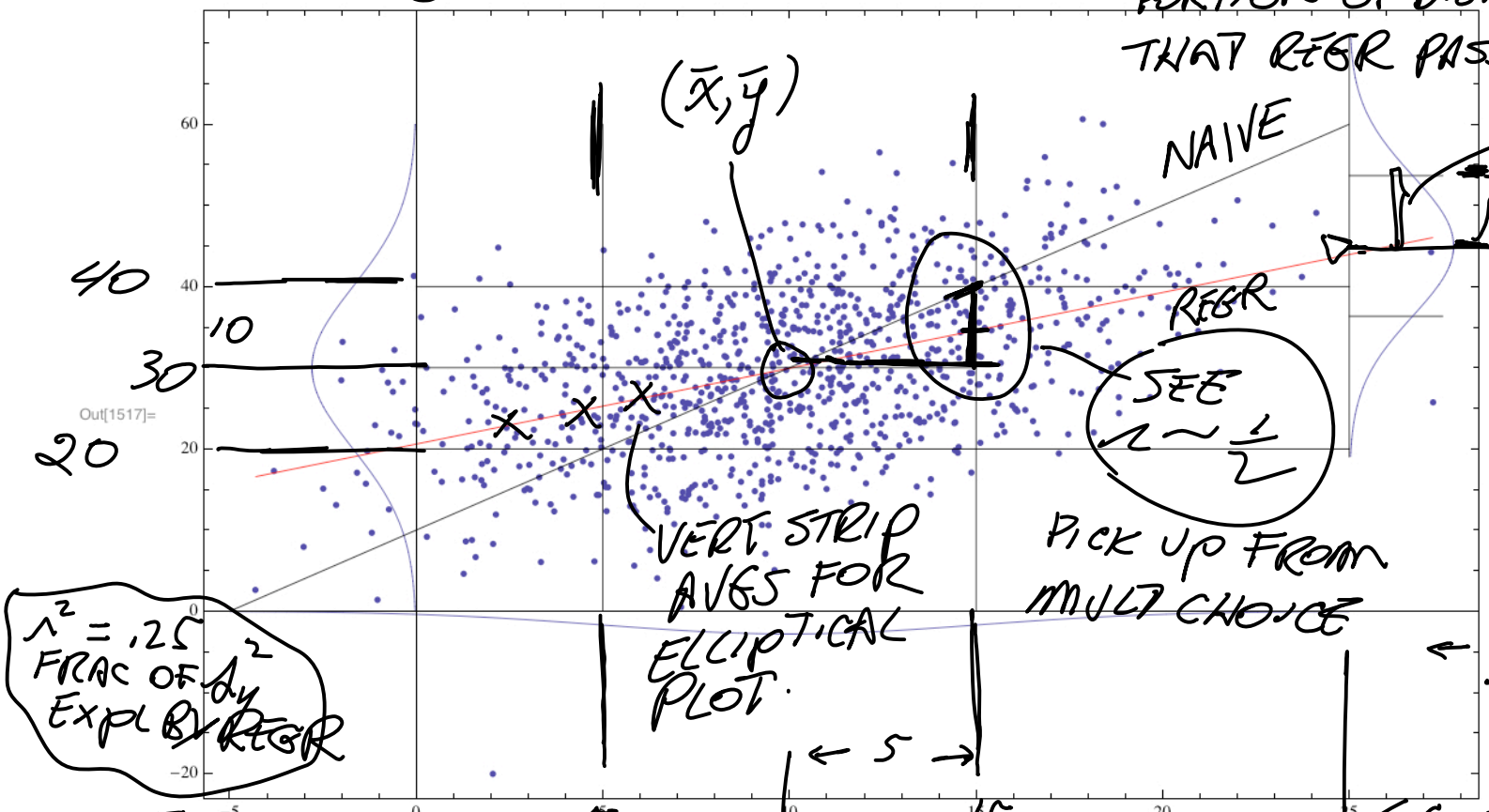


$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ $\left(\hat{\sigma}_x\right) = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ HAT = "TO ESTIMATE"

Stat 200 3-2-09

In[1517]:= regrpic[10, 30, 5, 10, 0.5, 1000, 70, .005]

Out[1517]=



$r = \frac{1}{2}$

r CAN BE SEEN AS THE PORTION OF DIST UP TO NAIVE THAT REGR PASSES THRU.

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

NAIVE

REGR

$\sqrt{1-r^2}\,\sigma_y$

$\bar{y}$ AVG HT OF SONS

40

10

30

20

SEE $r \sim \frac{1}{2}$

PICK UP FROM MULT CHOICE

VERT STRIP AVGS FOR ELLIPTICAL PLOT.

$\leftarrow 5 \rightarrow$

$r^2 = .25$ FRAC OF σ_y^2 EXPL BY REGR

N

R

$\bar{x}$

$r \sim .5$??

COULD BE ANY X

SD OF Y FOR SPECIFIED $x = 25$ (ANY)

$15\sqrt{1-r^2}\,\sigma_y$

$\sqrt{1-.25}\ 15$

$\sigma_y (\sim \sigma_y) = 10$

$\bar{y} (\sim \mu_y) = 30$

$\sigma_x \approx \frac{15-5}{2} = 5$

$\bar{x} (\sim \mu_x) = 10$

$x = 25$

HT OF FATHER

$\left(\text{SLOPE}_{\text{NAIVE}}\right) = \sigma_y/\sigma_x = \frac{10}{5} = 2$

$\left(\text{SLOPE}_{\text{REGR}}\right) = r\,\sigma_y/\sigma_x = r(2) = \frac{1}{2}(2) = 1$

FRAC OF σ_y^2 EXPLAINED BY REGR $r^2 = (.5819)$

$$\hat{\sigma}_x = \sqrt{\overline{x^2} - \bar{x}^2}$$

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

```
regrtable[{1, 4, 2, 7}, {2, 2, 8, 0}]
```

MatrixForm=

x	y	x^2	y^2	xy
1	2	1	4	2
4	2	16	4	8
2	8	4	64	16
7	0	49	0	0
—	—	—	—	—
3.5	3.	17.5	18.	6.5

COLUMN AVGS →

$\bar{x} = 3.5 \quad \bar{y} = 3$

$\overline{x^2} = 17.5 \quad \overline{y^2} = 18$

$\overline{xy} = 6.5$

$$\hat{\sigma}_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{17.5 - 3.5^2}$$

$$\hat{\sigma}_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{18 - 3^2} = 3$$

SLOPE NAIVE $\hat{\sigma}_y / \hat{\sigma}_x = \dfrac{3}{\sqrt{17.5 - 3.5^2}}$

SLOPE REGR $r\,\hat{\sigma}_y / \hat{\sigma}_x = .7619$

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{17.5 - 3.5^2}(3)} = \frac{6.5 - 3.5(3)}{\sqrt{17.5 - 3.5^2}(3)}$$

```
regrstats[{1, 4, 2, 7}, {2, 2, 8, 0}]
{3.5, 3., [redacted] -0.581914, -0.76190
```

$$(-0.581914)^2$$

$$0.338624$$

$$\sqrt{1 - .338624}\ 3.4641$$

$$2.81718$$

MCR MULTIPLE LINEAR REGR.

DATA	y	x_1	x_2	x_3	
	23	65	119	99	← SCORES OF SUBJECT 1

F.D MODEL $y = \underbrace{b_0}_{\text{CONST}} + b_1 x_1 + b_2 x_2 + b_3 x_3$

IF WE DO FIND EVIDENCE OF "GOOD FIT" ($R^2 \sim 1$)

[THEN REGR PREDICTOR $\hat{y}$ (FOR GIVEN $\underline{x}$) MIGHT BE RECOMMENDED.] MULT COR R

[IF PLOT IS ELLIPTICAL "GOOD NEWS" WE CAN STRONGLY RECOMMEND $\hat{y}$ — AND $\hat{y}$ WILL BE "CLOSE TO y" IF $R \sim 1$.]

EST'D COEFF

$\hat{y} = (\text{JOHN, SUBJ 1}) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3$

LET'S SUPPOSE YOU ARE TOLD

$\hat{b}_0 = .81 \quad \hat{b}_1 = 1.6 \quad \hat{b}_2 = 2.4 \quad \hat{b}_3 = -.61$

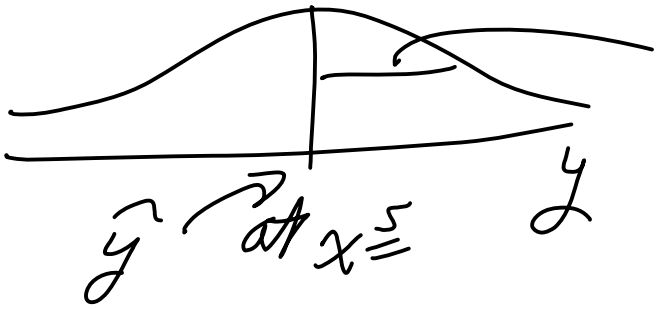
THEN FOR A SUBJECT WITH

$x_1 = 59 \quad x_2 = 94 \quad x_3 = 2.6$ (SAY)

$\hat{y} = \hat{b}_0 + \hat{b}_1(59) + \hat{b}_2(94) + \hat{b}_3(2.6)$ FEED x SCORES INTO COEFF ESTD BY C.S.

$= .81 + 1.6(59) + 2.4(94) - .61(2.6)$

PREDICTED (BY MLR) y SCORE FOR A SUBJECT WITH THE GIVEN x-SCORES.



$\sqrt{1-R^2}\,s_y$