SOME CALCULATIONS -

FOR A GIVEN DATASET, A COLLECTION OF (x,y) COORDINATE PAIRS $\{x_i, y_i\}_{i=1}^{n}$:

$i=$	$x_i=$	$y_i=$
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	
n	x_n	y_n

LET $\bar{x}$, s_x^2 BE THE MEAN AND VARIANCE FOR x

$\bar{x} = \frac{\sum x_i}{n}$, $s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$, SIMILARLY FOR $\bar{y}$, s_y^2.

THEN, THE DATASET IN STANDARD UNITS (SU) IS

i	$x_i(SU)$	$y_i(SU)$
1	$\frac{x_1 - \bar{x}}{s_x}$	$\frac{y_1 - \bar{y}}{s_y}$
2	$\frac{x_2 - \bar{x}}{s_x}$	$\frac{y_2 - \bar{y}}{s_y}$
$\vdots$		
n	$\frac{x_n - \bar{x}}{s_x}$	$\frac{y_n - \bar{y}}{s_y}$

AND WE DENOTE $\frac{x_i - \bar{x}}{s_x}$ BY $x_i(SU)$,

SIMILARLY FOR $y_i(SU)$

(AND, OF COURSE $s_x = \sqrt{s_x^2}$ AND $s_y = \sqrt{s_y^2}$)

WE DEFINE COVARIANCE OF x AND y TO BE

$$\sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

AND WE LET

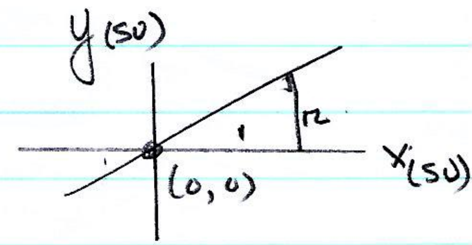
$$r = \frac{1}{s_x s_y} \cdot \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

CALLED CORRELATION COEFFICIENT

$$= \frac{1}{n} \sum \underbrace{\frac{(x_i - \bar{x})}{s_x}}_{x_{i\,(SU)}} \cdot \underbrace{\frac{(y_i - \bar{y})}{s_y}}_{y_{i\,(SU)}}$$

THE [LINEAR] "REGRESSION LINE TO THE MEAN" IS DEFINED TO BE

$$y_{(SU)} = r \; x_{(SU)}$$



WHICH CONTAINS $(0, 0)$ WITH SLOPE r

(I.E. WHEN x AND y ARE EXPRESSED IN STANDARD UNITS, THE REGRESSION LINE

- CONTAINS $(0, 0)$ AND HAS SLOPE r.
- r MEASURES "STRENGTH OF LINEAR RELATIONSHIP"
- r IS NOT SLOPE OF REGRESSION LINE FOR x AND y IN ORIGINAL UNITS

THE REGRESSION LINE IN ORIGINAL UNITS IS

$$\frac{y - \bar{y}}{s_y} = r \, \frac{x - \bar{x}}{s_x}$$

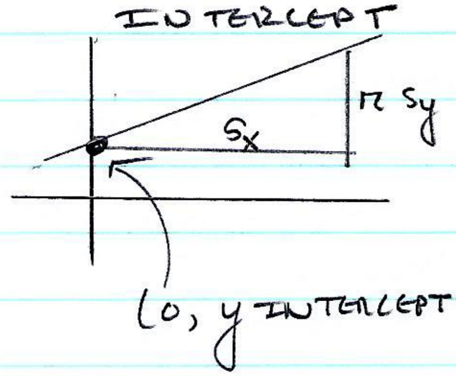
OR

$$y = r\left(\frac{x - \bar{x}}{s_x}\right) s_y + \bar{y}$$

$$= \underbrace{r \frac{s_y}{s_x}}_{\text{SLOPE}} \, x + \underbrace{\bar{y} - r \frac{s_y}{s_x} \bar{x}}_{\text{INTERCEPT}}$$

THINK
$y = mx + b$

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



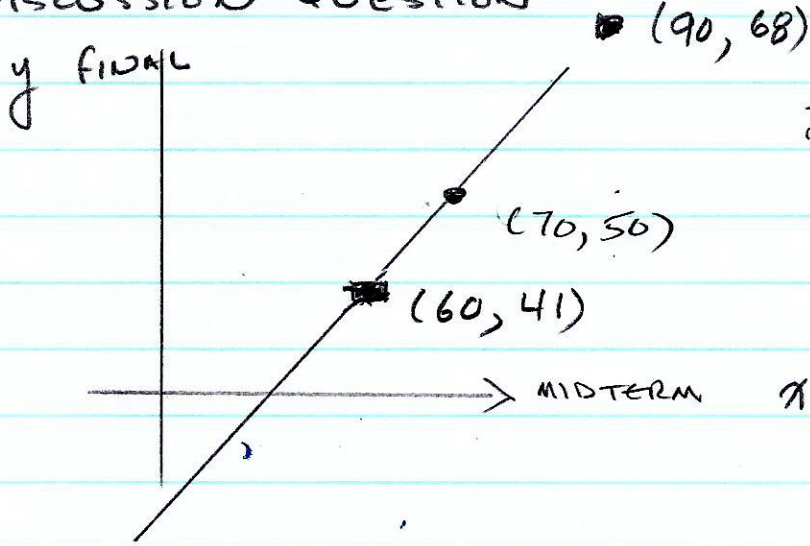
So, GIVEN x, WE "ESTIMATE y" BY COMPUTING y IN ABOVE EQUATION

THE MANNER IN WHICH VALUES FOR y ARE ACTUALLY DISTRIBUTED ABOUT "ESTIMATE FOR y GIVEN x" IS ANOTHER STORY.

THE APPROPRIATENESS OF THINKING ABOUT RELATIONSHIP BETWEEN x AND y AS "LINEAR" IS ANOTHER QUESTION.

(SEE lec 31, 0.0.1)

DISCUSSION QUESTION



$\bar{x} = 70 \qquad \bar{y} = 50$

$S_x = 10 \qquad S_y = 12$

$r = 0.75$

ESTIMATE FINAL EXAM FOR MIDTERM OF 90

$$y = r\frac{S_y}{S_x}x + \left(\bar{y} - r\frac{S_y}{S_x}\bar{x}\right)$$

$$= 0.75\ {}^{12}/_{10}\ x + 50 - 0.75\ {}^{12}/_{10} \cdot 70$$

$$= .9x + (50 - .9(70))$$

$$= .9x - 13$$

WHEN $x = 90$, $y = .9(90) - 13 = 68$

WHEN $x = 60$, $y = .9(60) - 13 = 41$

(SEE lec 31, 0.02)