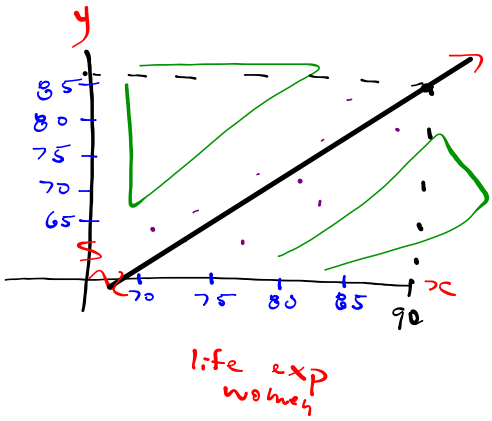
Unit 5: Regression Line

$y = mx + b$ (m: slope, b: y-int)

r: reliability of predictions

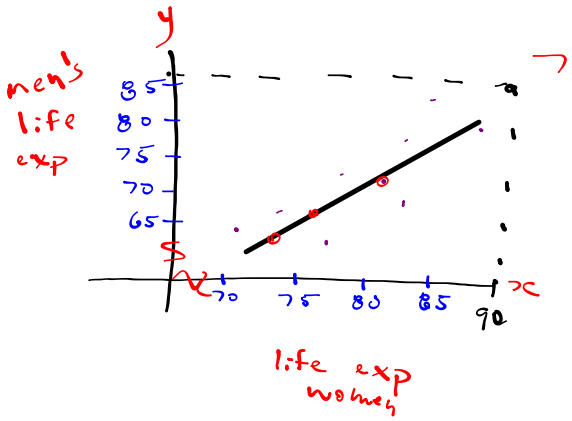


Definition: Regression Line

→ the line that best represents the linear relation/trend between 2 distributions

* we use $y = mx + b$ to make prediction about value of y when x is given, or vice versa

- Criteria for drawing Regression Line



- even amount of points on each side
- line goes through as many points as possible

2 graphically ways of finding $y = mx + b$

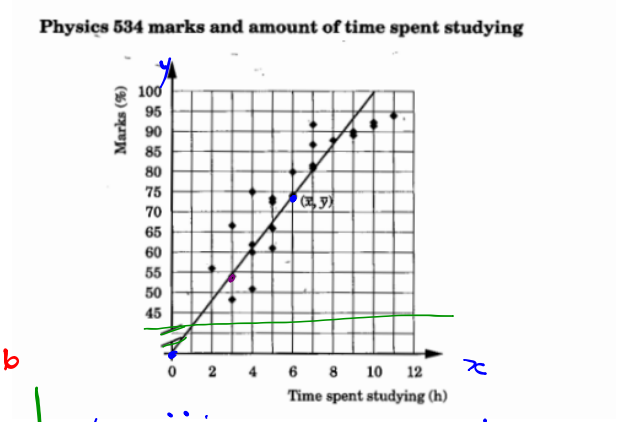
• mean Scatter Plot Method

P 5.4

step i.
Find $(\bar{x}, \bar{y})$
Plot it, and
Regression must
go thro it.
following criteria

P 4.20

(+time) x	(marks) y
3	54
	$\bar{y} = 74.46$



step ii
First find m
in $y = mx + b$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

use $(\bar{x}, \bar{y})$ $\bar{x} = 6$
and another (x, y) on line!

$(6, 74.46)$ $(3, 54)$
x_1 y_1 x_2 y_2

$$m = \frac{54 - 74.46}{3 - 6}$$

$$m = 6.82$$

step iii
find last parameter by
subbing a point $(\bar{x}, \bar{y})$

sub $(6, 74.46)$
x y

$y = 6.82x + b$

$74.46 = 6.82 \cdot 6 + b$

$74.46 = 40.92 + b$ (−40.92)

$b = 33.54$

$b =$

$$\boxed{y = 6.82x + 33.54}$$

Q: What mark will you get if you study for 13 hr?
x

$y = 6.82(13) + 33.54$
$y = 122\%$
↳ adjust for context
↳ remember still a prediction

A student studying for 13h should get a fairly perfect (100%) mark.

Using the mean scatter plot method, make a prediction about a man's life expectancy if the woman's life expectancy is 90 years old.

- use your graph and the 2 distri. on page. 3.29

2 graphically / algebraic way of finding Regression Line

→ Median – Median method.

p 4.20

(time) x	(marks) y

since 24 data points → 3 groups of 8

p 5.6

step i Sort the 2 distributions w respect to x into 3 groups (3 even groups or 1 of 3 w/ same group #)

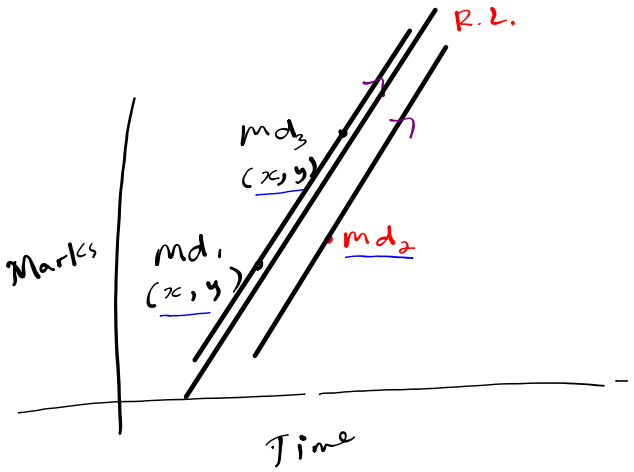
g1		g2		g3	
x	y	x	y	x	y

step ii Find the medians of the x-values in 3 groups. Find the medians of the y-values after you sort them in their individual groups.

$md_{x_1} = (3.5, 58)$ $md_{x_2} = (5.5, 73.5)$ $md_{x_3} = (9, 91)$

step iii find $y = mx + b$ of line md_1, md_3

$m = \frac{y_2 - y_1}{x_2 - x_1}$ $md_1 (3.5, 58)$ $md_3 (9, 91)$



$m = 6$

$y = 6x + b$

$b = 37$

$l_{md_1 md_3}$: $y = 6x + 37$

step iv: Find $y = mx + b$ of $l\, md_2$

since $l_{md_1 md_3} \parallel l_{md_2}$

then $m_{1/3} = m_2$

$m_2 = 6$

$md_2 (5.5, 73.5)$

$y = 6x + b$

$b = 40.5$

l_{md_2} : $y = 6x + 40.5$

step v find $y = mx + b$ of regression line. Same slope and y-int is the weighted average of $b_{1/3}$ and b_2

$m_{RL} = b$ $b_{RL} = \frac{37 + 37 + 40.5}{3}$

$b_{RL} = 38.16$

l_{RL} : $y = 6x + 38.16$

p 5.10 → calculator instructions

RCL b → slope

RCL ^ → y-int

Using the median – median method, make a prediction about a man's life expectancy if the woman's life expectancy is 90 years old.

HMWK p 5.14 – 5.21 #1/#2 not 1.b) not 2b)

use the 2 distributions on page. 3.29

88% 6-credits — 3 cr 88

82% 3-credits — 82

$$\frac{2(88) + 82}{3}$$