Homework for Regression

Use StatCenter's StatTool to analyze the data in problems 2 and 5 of this homework. The data is listed as Regression_2.dat and Regression_5.dat in the dialogue box that appears when you use the drop-down menus under "File," then "Open from File." You can check all the relevant answers for these problems using StatTool.

1. Graph the following linear equations:
   a. $Y = 2X + 6$
   b. $Y = .5X - 4$
   c. $Y = -1.5X + 10$
   d. $Y = X$
   e. $Y = -2X$
   f. $Y = .5X$
   g. $Y = X + 1$
   h. Make general summary of how the parameter "a" (Y intercept) worked in the above equations.
   i. Make a general summary of how the parameter "b" (slope) worked in the above equations.

2. Find the equation for the regression line of Y on X for the following data.

X	1	2	3	4	5
Y	5	4	2	3	1

3. In a study concerned with the relationship between two variables, X (High School GPA) and Y (College GPA), the following was obtained:

High School GPA	College GPA
$\bar{X} = 2.20$	$\bar{Y} = 2.75$
$s_x = 0.80$	$s_y = 0.55$

$r = 0.70$
$N = 100$

   a. Write the prediction equation for predicting College GPA from High School GPA
   b. Sally B. obtained a score of 3.20 on the X-variable (High School GPA). Predict her score on the Y-variable (College GPA).

4. Assume that $M_x = 10$, $s_x = 5$, $M_y = 45$, $s_y = 8$. Draw a separate graph for each regression line for the following values of r.
   a. 0.00 b. 0.20 c. 0.40 d. 0.60 e. 0.80 f. 1.00
   Generalize: What is the relationship between the size of r and the angle formed by the regression lines? If the values of r given in (b) through (f) above were all negative, what is the relationship?

5. The scores of ten students on both a midterm and final exam in a statistics class are recorded the table below.

Student #	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Midterm	30	12	15	25	22	28	18	16	24	28
Final	27	13	10	17	19	25	20	12	22	26

A) Draw a scatterplot

B) Find the correlation (r) between Midterm and Final.

C) Find the regression line for predicting Final scores from Midterms. Draw this on your scatterplot.

D) Calculate the Total Variance in the Final Exam scores. Also calculate the Explained and Unexplained Variance in Final Exam scores. Use r^2.

E) Represent the Total, Explained, and Unexplained Y variance in a circular drawing like we did in lecture.

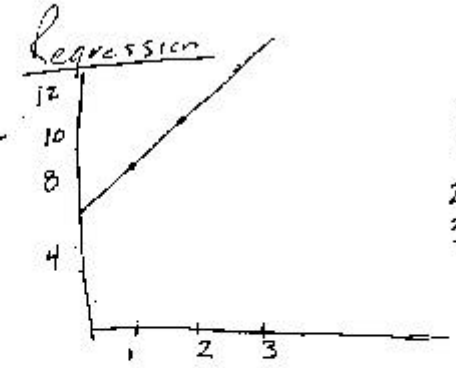
F) Calculate the proportion of variance accounted for. This is sometimes called the coefficient of determination. Also calculate the proportion of variance not accounted for.

G) If a student scored 17 on the midterm, what would you predict that student's score would be on the final exam?

I)Calculate the Prediction Error Variance (the long way, not using r). That is, fill out the following table and then calculate the variance of the errors.

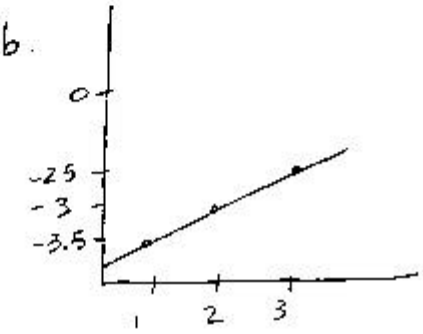
Sub #	Mid	Fin	Y'	e	e^2
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

(1)

a. 

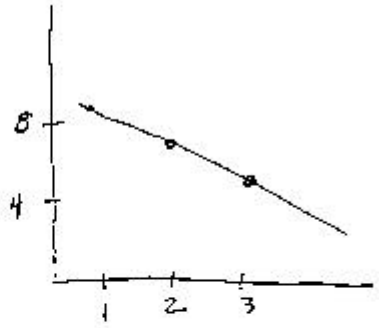
X	Y
1	8
2	10
3	12

$Y = 2X + 6$

b. 

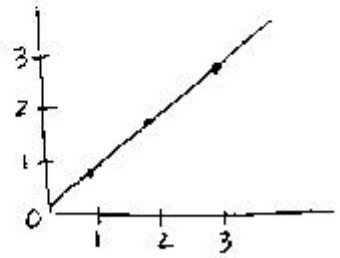
X	Y
1	-3.5
2	-3
3	-2.5

$Y = .5X - 4$

c. 

X	Y
1	8.5
2	7
3	5.5

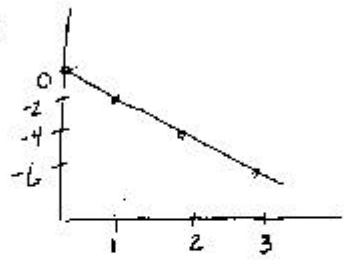
$Y = -1.5X + 10$

d. 

X	Y
1	1
2	2
3	3

$Y = X$

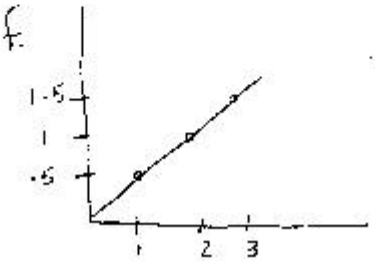
e



X	Y
1	-2
2	-4
3	-6

$Y = -2X$

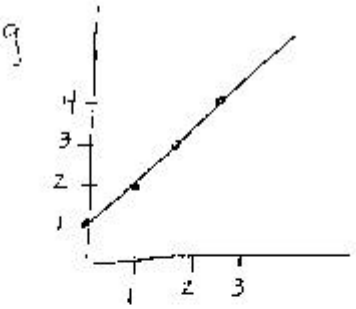
f.



X	Y
1	.5
2	1
3	1.5

$Y = .5X$

g



X	Y
1	2
2	3
3	4

$Y = X + 1$

h Y intercept "a": Where the line intercepts the Y-axis (The value of Y, when X=0)

i Slope "b": Determines if line slopes up or down & also the steepness of the line

2

X	Y
1	5
2	4
3	3
4	2
5	1
$M = \frac{\Sigma X}{n} = \frac{15}{5} = 3$	$3 = M$

$r_{xy} = -1$

$b = r_{xy}\left(\frac{S_y}{S_x}\right)$

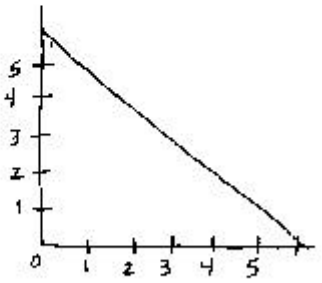
$= (-1)\left(\frac{1.414}{1.414}\right)$

$= -1$

$a = M_y - b_y M_x$

$3 - (-1)(3)$

$= 6$



$S_x^2 = \frac{\Sigma X^2}{n} - M^2$

$= \frac{1+4+9+16+25}{5} - 9$

$= 11 - 9$

$= 2$

$S_x = \sqrt{2} = 1.414$ $S_y = \sqrt{2} = 1.414$

$Y = a + bX$

$\boxed{Y = 6 - X}$

3 $b = r_{xy}\left(\frac{S_y}{S_x}\right)$

$= (.7)\left(\frac{.56}{.80}\right)$

$= .48$

$a = M_y - b_y M_x$

$= 2.75 - (.48)(2.20)$

$= 1.694$

$Y = a + bX$

$Y = 1.694 + .48X$

b $\hat{Y} = 1.694 + .48(3.2)$

$= 3.23$

$M_x = 10 \quad S_x = 5 \quad M_y = 45 \quad S_y = 8$

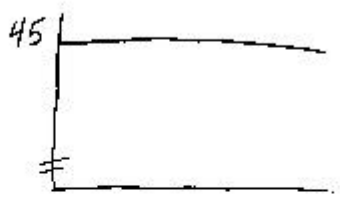
$r_{xy} = 0$ a. $b = r_{xy}\left(\frac{S_y}{S_x}\right)$ $\quad a = M_y - bM_x$

$= 0\left(\frac{8}{5}\right)$ $\quad = 45 - (0)(10)$

$= 0$ $\quad = 45$

$Y' = 45 + 0X$

or $Y' = 45$



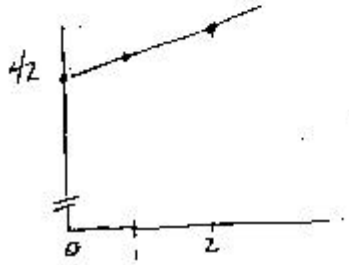
$r_{xy} = .20$ b. $b = r_{xy}\left(\frac{S_y}{S_x}\right)$ $\quad a = 45 - (.32)(10)$

$= .20\left(\frac{8}{5}\right)$ $\quad = 41.8$

$= .32$

$Y' = 41.8 + .32X$

X	Y
0	41.8
1	42.12
2	42.44

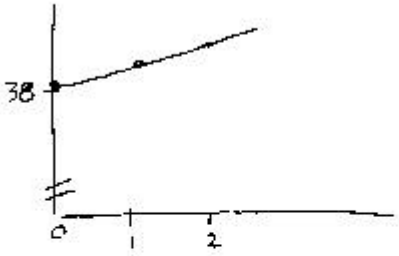


$r_{xy} = .40$ c. $b = (.40)\left(\frac{8}{5}\right)$ $\quad a = 45 - (.64)(10)$

$= .64$ $\quad = 38.6$

$Y' = 38.6 + .64X$

X	Y
0	38.6
1	39.24
2	39.88

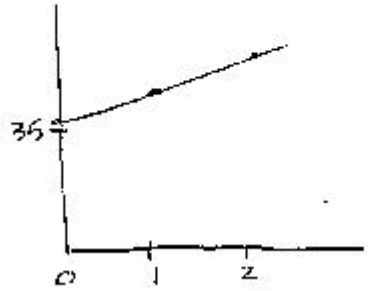


$r_{xy} = .60$ d. $b = (.60)\left(\frac{8}{5}\right)$ $\quad a = 45 - (.96)(10)$

$= .96$ $\quad = 35.4$

$Y' = 35.4 + .96X$

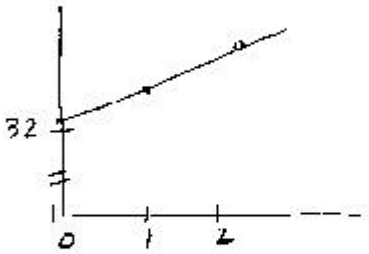
X	Y
0	35.4
1	36.36
2	37.32



$r_{xy} = .80$ e $b = (.80)\left(\frac{8}{5}\right) = 1.28$ $a = 45 - (1.28)(10) = 32.2$

$\boxed{Y' = 32.2 + 1.28X}$

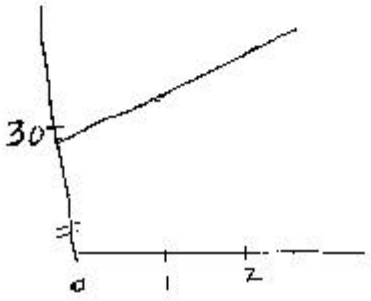
X	Y
0	32.2
1	33.48
2	34.76



$r_{xy} = 1.0$ f. $b = (1)\left(\frac{8}{5}\right) = 1.6$ $a = 45 - (1.6)(10) = 29$

$\boxed{Y' = 29 + 1.6X}$

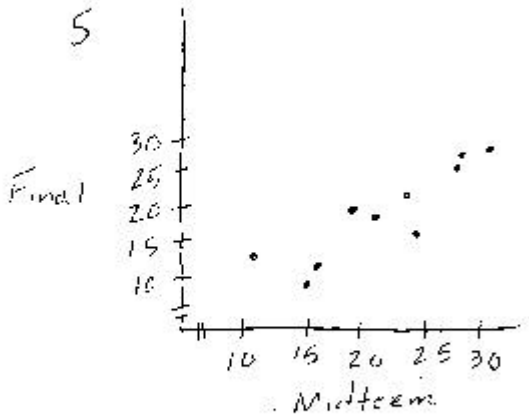
X	Y
0	29
1	30.6
2	32.2



Generalize:
Size of r & angle: The larger the r, the steeper the angle
If b thru f were all negative, the regression line would slope downward.

Hw Regression Question #5

5



Mid	Fin
30	27
12	13
15	10
25	17
22	19
28	25
18	20
16	12
24	22
28	26

b $r_{xy} = \dfrac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma Y_i)}{\sqrt{(n\Sigma x^2 - (\Sigma x_i)^2)(n\Sigma Y^2 (\Sigma Y)^2)}}$

Mid (X)	X^2	Final (Y)	Y^2	XY
30	900	27	729	810
12	144	13	169	156
15	225	10	100	150
25	625	17	289	425
22	484	19	361	418
28	784	25	625	700
18	324	20	400	360
16	256	12	144	192
24	576	22	484	528
28	784	26	676	728
218	5102	191	3977	4467

$r_{xy} = \dfrac{(10)(4467) - (218)(191)}{\sqrt{(10)(5102) - 218^2)((10)(3977) - 19}}$

$= \dfrac{44670 - 41638}{\sqrt{(51,020 - 47,524)(39770 - 36}}$

$= \dfrac{3032}{\sqrt{(3496)(32841)}}$

$= \dfrac{3032}{3390}$

$= .894$

c. $b = r_{xy}\left(\frac{S_y}{S_x}\right)$

$= .894\left(\frac{5.736}{5.913}\right)$

$= .867$

$a = M_y - bM_x$

$= 19.1 - (.867)(21.8)$

$= .199$

$S_x^2 = \frac{\Sigma X^2}{n} - M_x^2$

$= \frac{5102}{10} - \left(\frac{218}{10}\right)^2$

$= 510.2 - 475.24$

$= 34.96$

$S_x = \sqrt{34.96}$

$= 5.913$

$S_y^2 = \frac{\Sigma y^2}{n} - M_y^2$

$= \frac{3977}{10} - \left(\frac{191}{10}\right)^2$

$= 397.7 - 364.8$

$= 32.9$

$S_y = \sqrt{32.9}$

$= 5.736$

$Y' = a + bx$

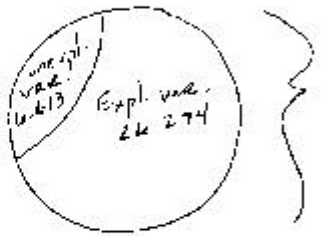
$= .199 + .867X$

d. Total variance exam scores $= S_y^2 = 32.9$

Explained variance $= r^2 S_y^2 = (.894)^2(32.9) = 26.294$

Unexplained variance $= (1 - r^2) S_y^2 = (1 - .799)(32.9) = 6.613$

e.



Total variance = 32.9

f. Proportion of explained variance = $\frac{26.294}{32.9} = .79$
Proportion of unexplained variance: $\frac{6.413}{32.9} = .21$

g. $Y' = .199 + .867X$
$= .199 + (.867)(17)$
$= 14.94$

i.

X	Y	Y'	e	e^2
30	27	26.21	.79	.62
12	13	10.6	2.40	5.76
15	10	13.2	<3.27>	10.24
25	17	21.67	<4.87>	23.72
22	19	21.33	<2.33>	5.43
28	25	24.48	.52	.27
19	20	15.81	4.19	17.56
16	12	14.07	<2.07>	4.28
24	22	21.01	.99	.98
28	26	24.48	1.52	2.31
				71.17

Predicted error variance = $\frac{\Sigma e^2}{n} - M_e^2$

$= \frac{71.17}{10} - 0^2$

$= 7.117 \approx 6.613$

Rounding Difference