

REGRESSION AND PREDICTION

LINEAR FUNCTIONS

$$Y = a + bX$$

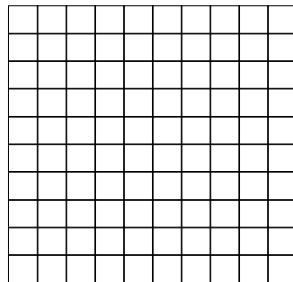
$$a =$$

$$b =$$

Examples

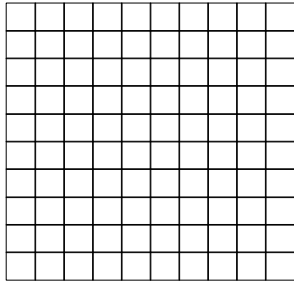
$$Y = 1 + 2x$$

X	Y



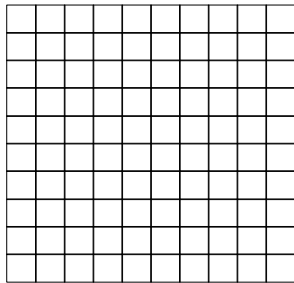
$Y = -1 + .5X$

X	Y



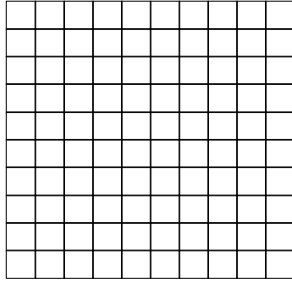
$Y = 2 - 1.5X$

X	Y



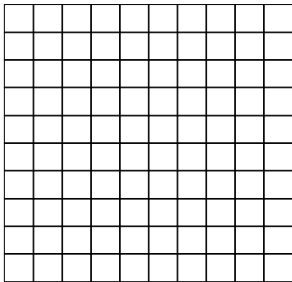
$Y = 3 + 0X$

X	Y



$Y = 0 + 1X$

X	Y



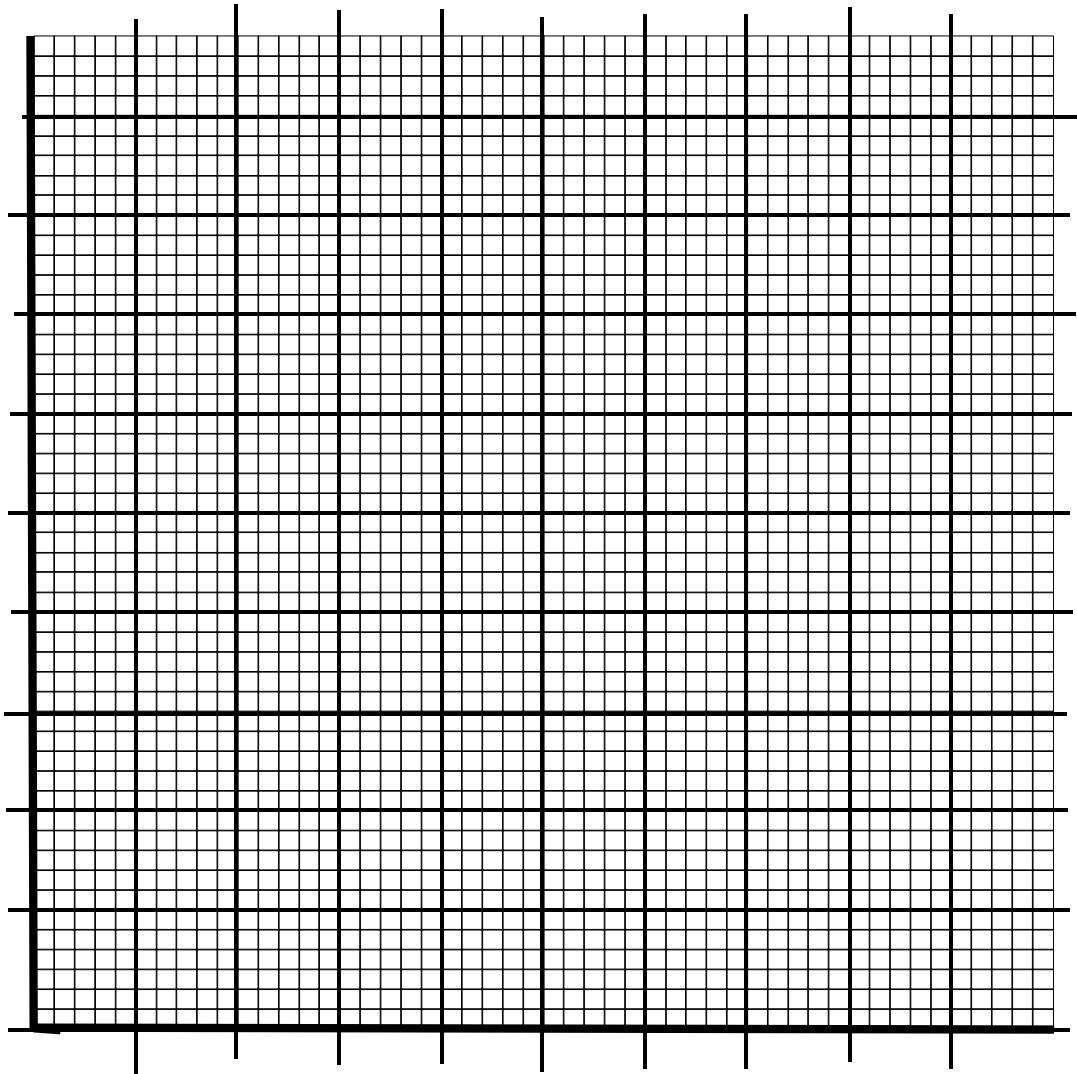
BEST FIT REGRESSION LINE**PREDICTOR VARIABLES AND CRITERION VARIABLES****EXAMPLE:**

Suppose we are interested in the relationship between # of Health Problems between the ages of 65 and 70 and the average number of packs of cigarettes smoked per day. We hypothesize that the number of packs will be positively related to the number of health problems. In fact we think we can predict to a certain degree the number of health problems from the number packs smoked. We randomly sample six people and obtain the following data:

<i>Person #</i>	<i># of packs per day (X)</i>	<i># of Health Problems (Y)</i>	X^2	Y^2	XY
1	3	6	9	36	18
2	10	23	100	529	230
3	6	16	36	256	96
4	1	3			
5	2	10			
6	8	8			
Sums:	30	66	214	994	431

$$S_x = 3.26598; S_y = 6.68331; M_x = 5; M_y = 11; r = .77119 \quad a = 3.109; b_y = 1.578$$

Scatterplot:



FORMULAS FOR a AND b (WHEN PREDICTING Y FROM X)

$$Y' = a_y + b_y X$$

 b_y a_y **FORMULAS FOR a AND b (WHEN PREDICTING X FROM Y)**

$$X' = a_x + b_x Y$$

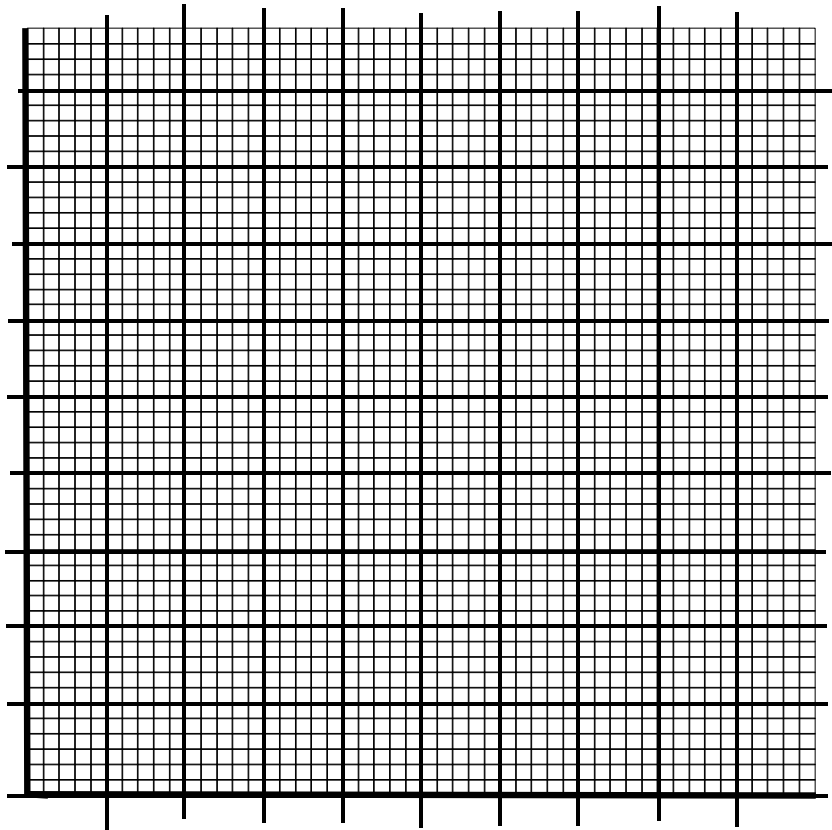
 b_x a_x **EXAMPLE CALCULATIONS OF REGRESSION FORMULAS:** $a_y =$ $b_y =$

REGRESSION EQUATION FOR PREDICTING Y from X:

GO BACK AND DRAW THIS REGRESSION LINE ON YOUR SCATTERPLOT

PRINCIPLE OF LEAST SQUARED ERROR

REDRAW THE SCATTERPLOT



X Y Y' e

Error

e =

For Subject #2, whose X value equals 10

Sum of e =

Principle of Least squared error:

The equations found by the formulas we are using has the least...

ERROR VARIANCE (Find the Variance of the prediction errors)

X	Y	Y'	e	e ²
1	3	4.687	-1.69	
2	10			
3	6			
6	16			
8	8			
10	23			

Mean of error =

Sum of error squared =

What is formula for Variance?

What is formula for Error Variance?

Calculate Error Variance

What does Error Variance Mean?

Error Variance is called Unexplained Variance in the next section

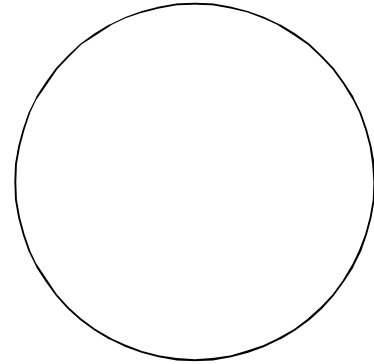
EXPLAINED VARIANCE AND UNEXPLAINED (ERROR) VARIANCE

Example: X =

Y =

$S_x = 3.26598$; $S_y = 6.68331$; $M_x = 5$; $M_y = 11$; $r = .77119$ $a = 3.109$; $b_y = 1.578$

The idea of variability in health problems (Y)



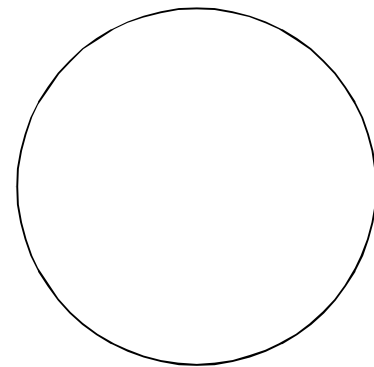
Total Variance of Y
 $S_y^2 = 44.66667$

Explained Variance

$$= r^2(\text{Total Variance}) = r^2 S_y^2$$

Unexplained Variance

$$= (1 - r^2) S_y^2$$



Unexplained = Error Variance

PROPORTION OF VARIANCE EXPLAINED (ACCOUNTED FOR)

Proportion of Var Explained =

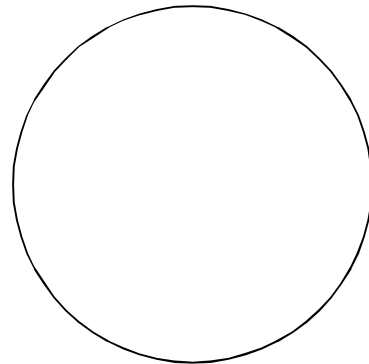
$$r^2 =$$

r^2 is sometimes called the **coefficient of determination**

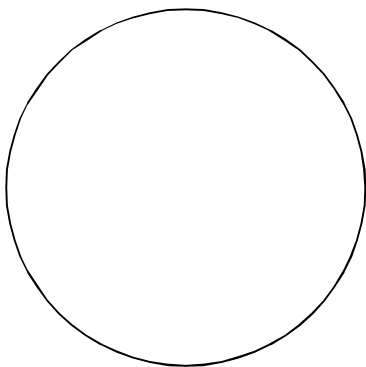
Proportion of Variance Unexplained (NOT accounted for)

Proportion of Var Unexplained =

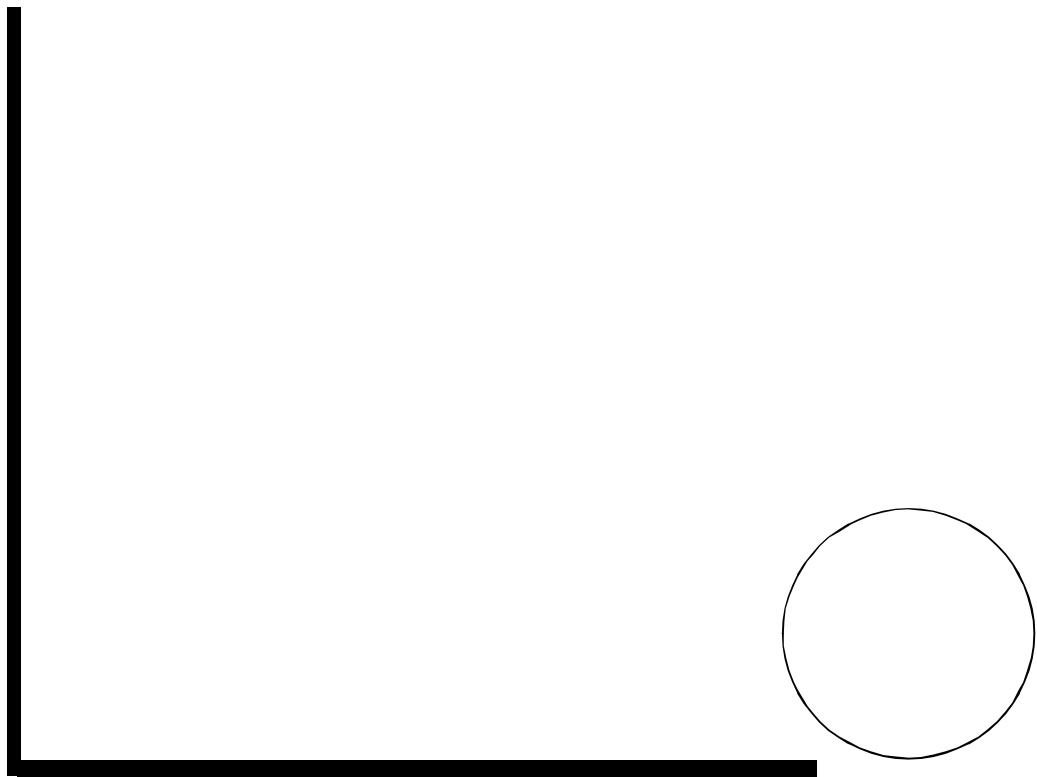
$$1 - r^2 =$$



Expressing the above ideas as percentages



Integration of Explained and Unexplained Variance with graph and Y' line



The number health problems in later life varies from person to person. Some of this total variance in health problems can be accounted for (explained by) cigarette smoking and some of the total variance in health problems cannot be explained by cigarettes.

Summary and overview of Regression

Linear Functions in algebra

Estimating the parameters (a and b) of a linear function from data

Describing the relationship between two variables as a straight line

Total Variance in the Criterion Variable (Y)

Explained Variance

Unexplained Variance (Error Variance)