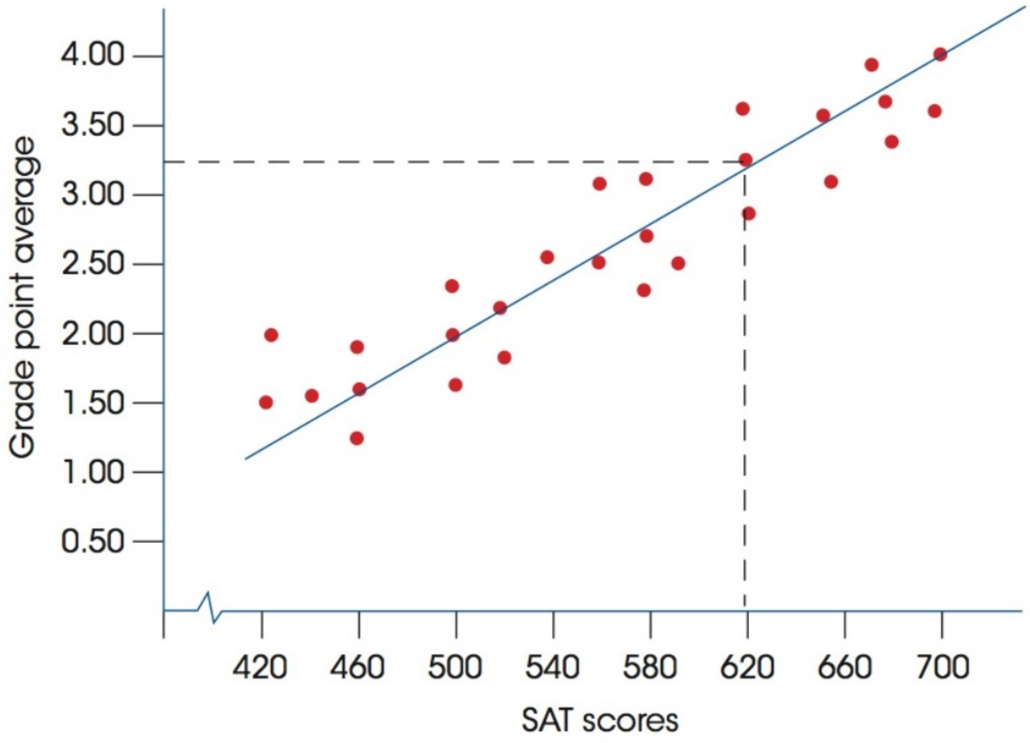
MGMT 9050: Quantitative Methods I

Linear Regression

Wilbur O. and Ann Powers College of BUSINESS

THE CONTEXT



Grade point average
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
420
460
500
540
580
620
660
700
SAT scores

DEFINING A LINE

$$Y = B_0 + B_1 X$$

- B_0 indicates Y when X is 0.
- B_1 is the slope of the line.

USING A LINE

$$Y = 35 + 15X$$

- If X = 3, Y = 80
- If X = 8, Y = 155
- If X = -2, Y = 5

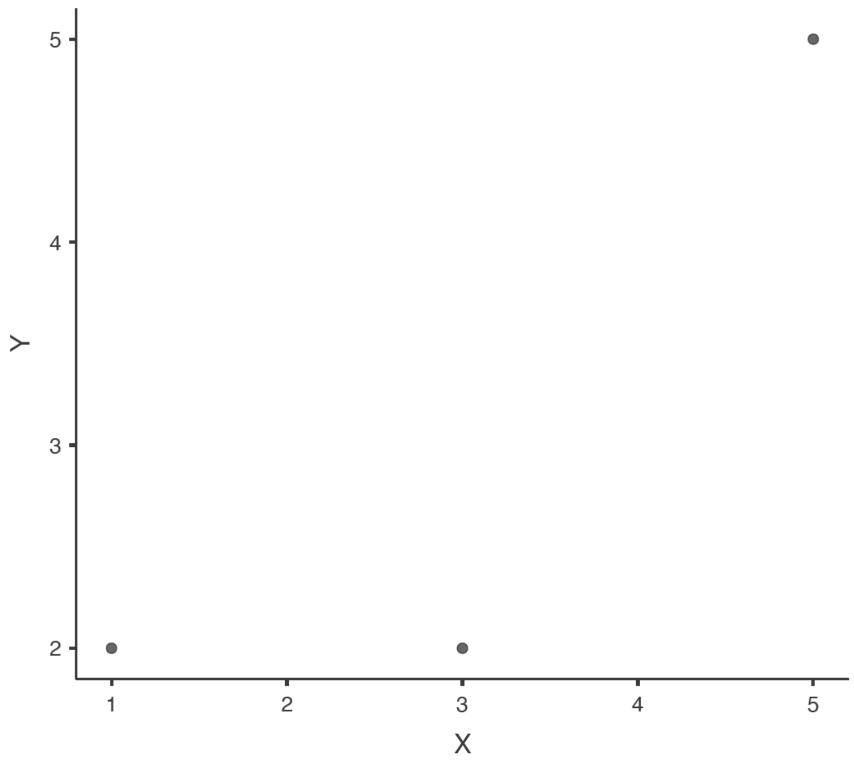
WHAT'S THIS MEAN FOR US?

- We don't start with an equation, but rather data
- We must construct the equation from the data
- Our data never fall cleanly along a straight line, so:

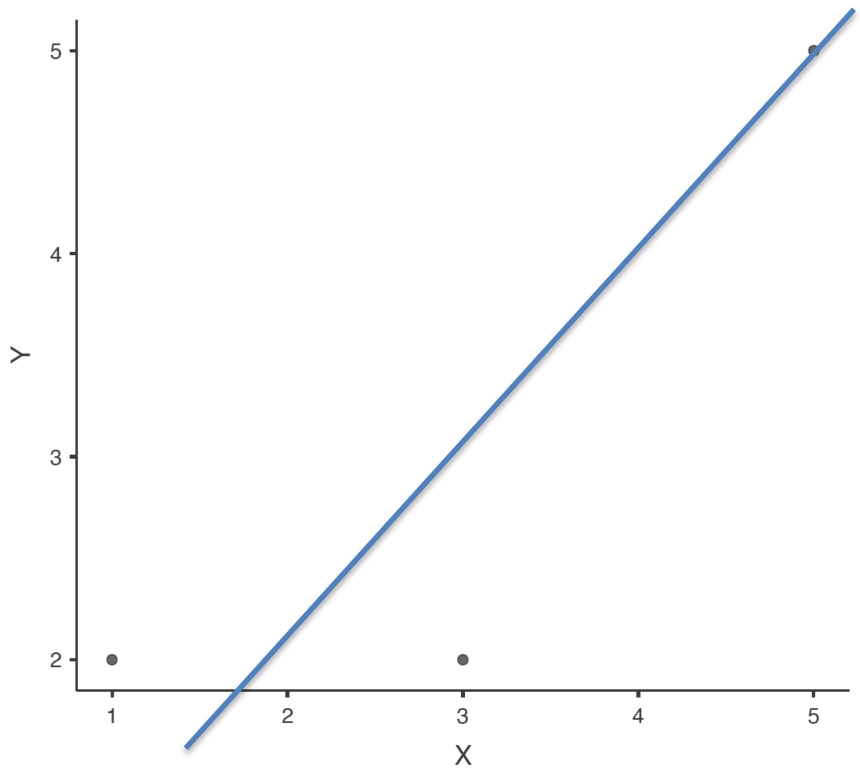
$$Y = B_0 + B_1 X + e$$

$$\hat{Y} = B_0 + B_1 X$$

EXAMPLE

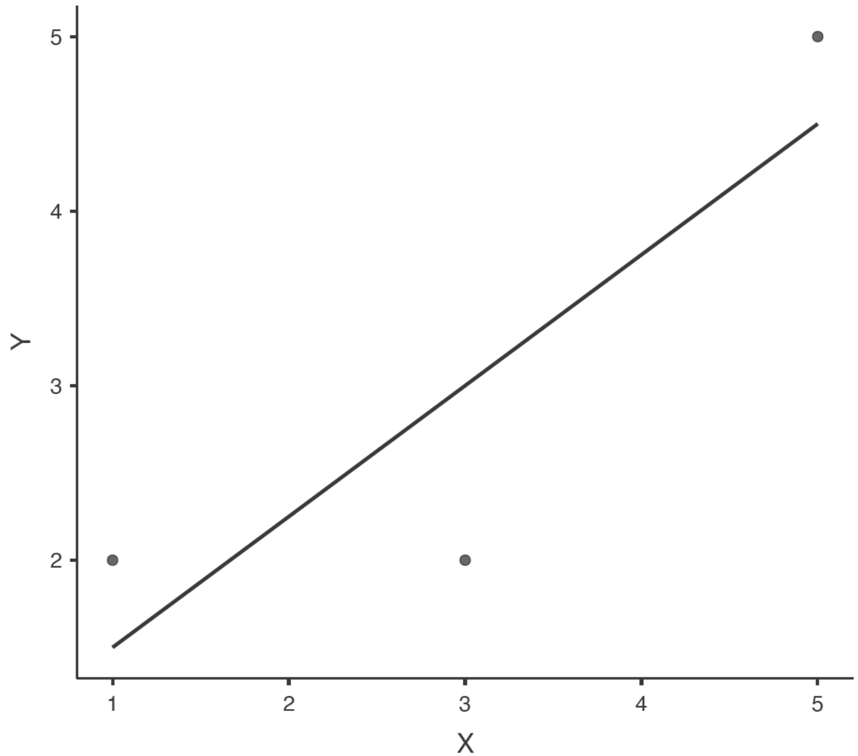


A POSSIBLE LINE



$$Y = 0 + 1X$$

LINE OF BEST FIT



$$Y = .75 + .75X$$

COMPARE THE LINES

Model 1

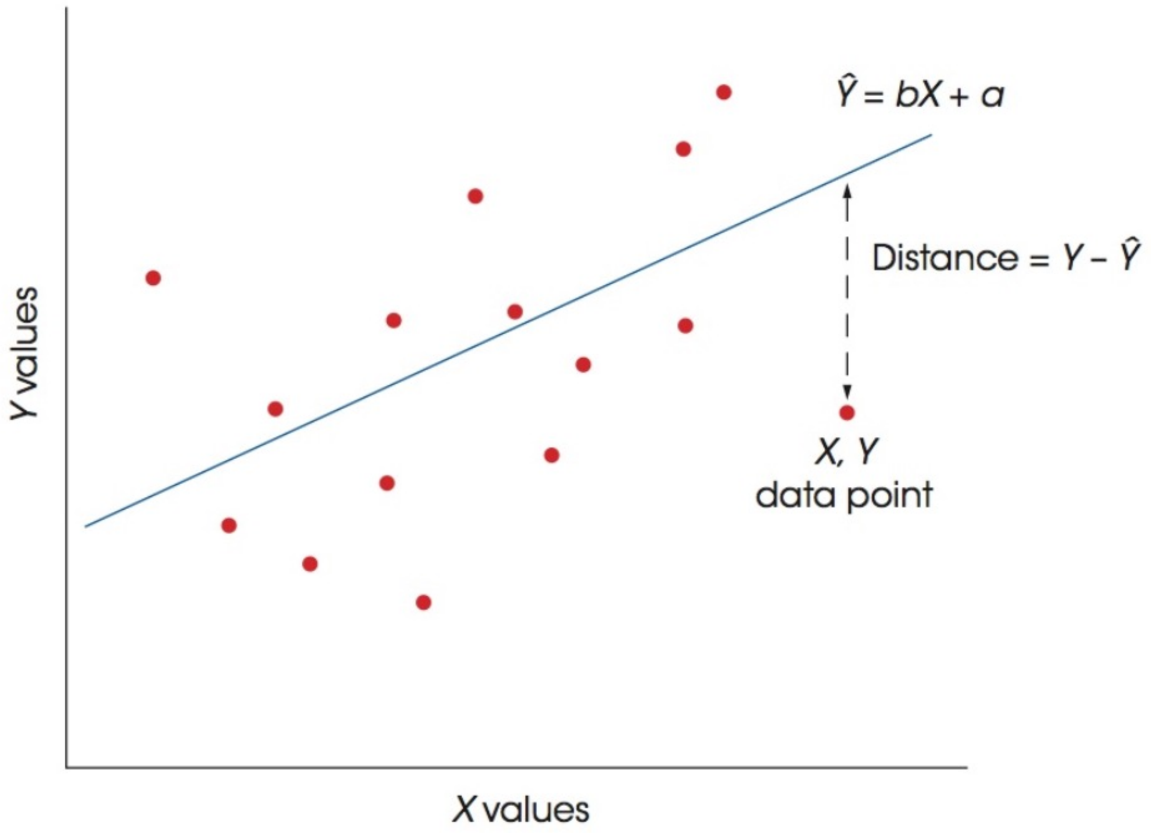
X	Y	Yhat	Y - Yhat	Squares
1	2	1	1	1
3	2	3	-1	1
5	5	5	0	0

SS = 2

Model 2

X	Y	Yhat	Y - Yhat	Squares
1	2	1.5	.5	.25
3	2	3	-1	1
5	5	4.5	.5	.25

SS = 1.5


$\hat{Y} = bX + a$
Distance = $Y - \hat{Y}$
X, Y
data point
Y values
X values

STANDARDIZED FORM

$$z_Y = \beta_1 z_X + e$$

$$\hat{z}_Y = \beta_1 z_X$$

Wilbur O. and Ann Powers College of BUSINESS

SIMPLE DATA

Case	X	Y
A	5	10
B	1	4
C	4	5
D	7	11
E	6	15
F	4	6
G	3	5
H	2	0
I	4	8
J	8	12

OUTPUT

Model Fit Measures

				Overall Model Test			
Model	R	R^2	Adjusted R^2	F	df1	df2	p
1	0.846	0.716	0.681	20.182	1	8	0.002

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
B	127.758	1	127.758	20.182	0.002
Residuals	50.642	8	6.330		

Note. Type 3 sum of squares

Model Coefficients - C

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	-0.038	1.877	-0.020	0.984	
B	1.736	0.386	4.492	0.002	0.846

MODEL FIT

- Total Sum of Squares

$$SS_{TOT} = \sum (Y - \bar{Y})^2$$

- Regression Sum of Squares

$$SS_{REG} = \sum (\hat{Y} - \bar{Y})^2$$

- Residual Sum of Squares

$$SS_{RES} = \sum (Y - \hat{Y})^2$$

MODEL FIT

- R — correlation between Y and predicted Y
- R2 — percentage of variance in Y that can be explained given X
- F — $$F = \frac{MS_{reg}}{MS_{res}}$$
  - degrees of freedom
  - numerator = # of predictor variables (*k*)
  - denominator = *N - k - 1*

MODEL FIT

- Total Sum of Squares = 178.40
- Regression Sum of Squares = 127.76
- Residual Sum of Squares = 50.64
- R and R-Square = .85 & .72
- $F_{(1,8)} = 20.18$, $p = .002$