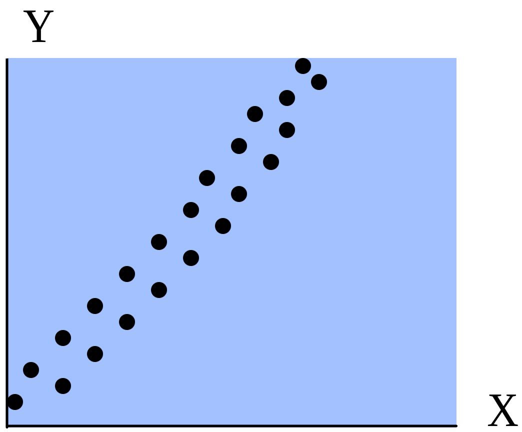
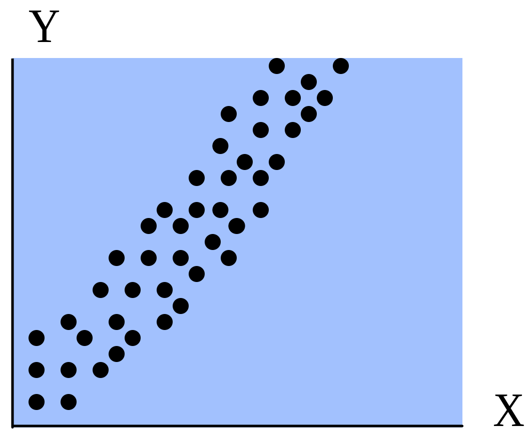
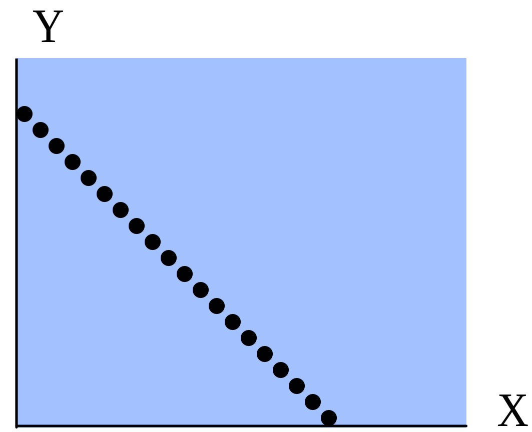
Types of Relationships Found in Scatter Diagrams



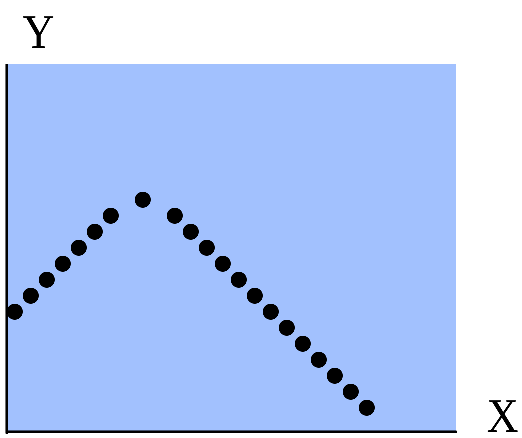
A--Strong Positive Linear Relationship



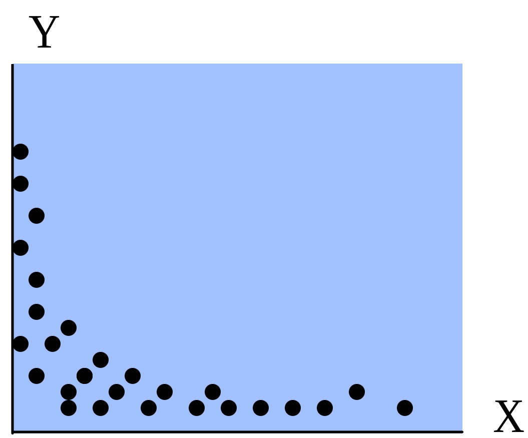
B--Positive Linear Relationship



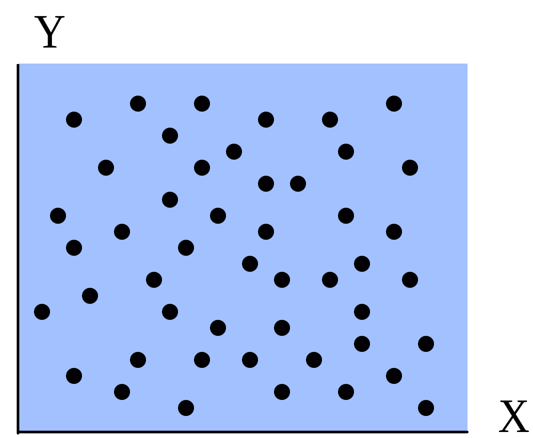
C--Perfect Negative Linear Relationship



D--Perfect Parabolic Relationship

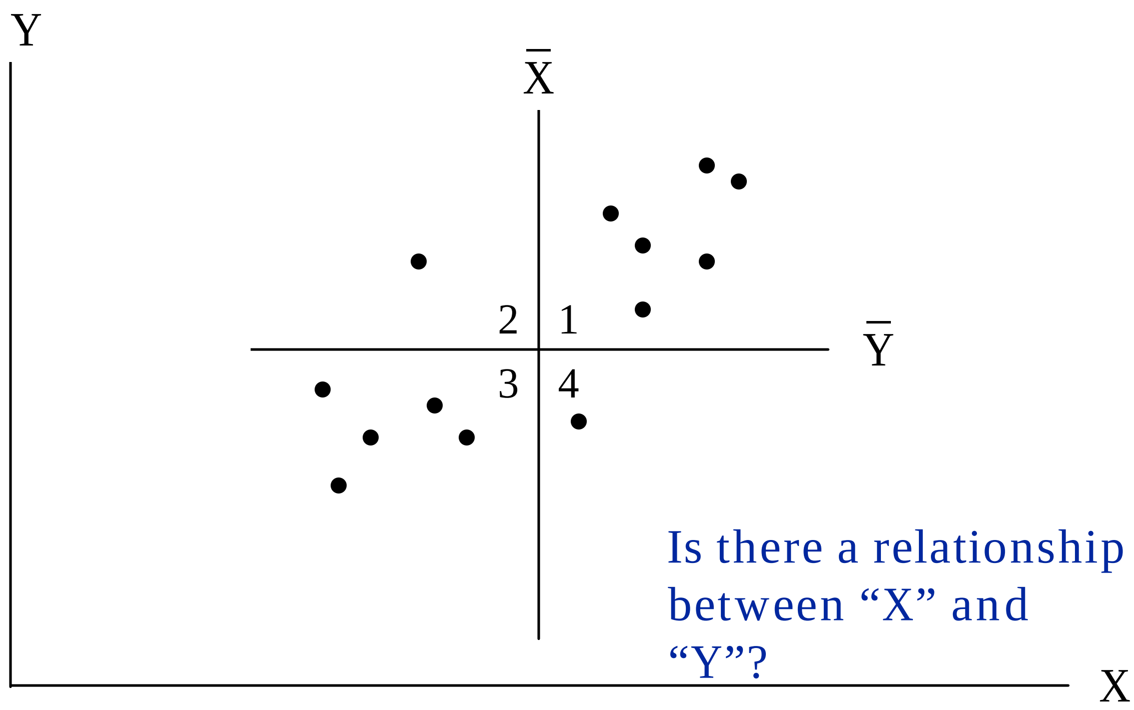


E--Negative Curvillinear Relationship



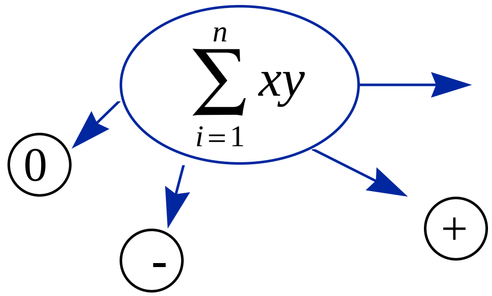
F--NoRelationship between X and Y

Product Moment Correlation



An example of a scatter diagram and associated quadrants.

$x = (X_i - \bar{X})$
$y = (Y_i - \bar{Y})$
$xy = (X_i - \bar{X})\,(Y_i - \bar{Y})$



Measure of the extent to which “X” and “Y” are associated <u>or</u> covary

$$\sum_{i=1}^{n} xy \quad = \quad f\,(\,n, \text{units of measurement})$$

these effects need to be eliminated

COVARIANCE BETWEEN X & Y $= \dfrac{;_{xy}}{n - 1} = \dfrac{;\,(X_i - \bar{X})\,(Y_i - \bar{Y})}{n - 1}$

CORRELATION COEFFICIENT *(“ $_{xy}$ “)* $= \dfrac{;_{xy}}{(n - 1)\,(S_x)\,(S_y)}$

$$xy \quad = \quad \frac{;\,(X_i - \bar{X})\,(Y_i - \bar{Y})}{\sqrt{;\,(X_i - \bar{X})^2 \;\; ;\,(Y_i - \bar{Y})^2}}$$

MEASURE OF (ASSOCIATION) $xy \longrightarrow$ + 1.00 to -1.00

$^{2}_{xy}$ = **PERCENTAGE OF VARIATION SHARED BY THE TWO VARIABLES**

COEFFICIENT OF DETERMINATION

MARKETING MANAGEMENT (*X*) AND MARKETING RESEARCH (*Y*) GRADES FOR 10 STUDENTS

X	Y	$x=(X-\bar{X})$	$y=(Y-\bar{Y})$	xy	x^2	y^2
75	85	1	10	10	1	100
80	85	6	10	60	36	100
60	65	-14	-10	140	196	100
55	60	-19	-15	285	361	225
85	80	11	5	55	121	25
95	95	21	20	420	441	400
70	60	-4	-15	60	16	225
75	80	1	5	5	1	25
80	80	6	5	30	36	25
65	60	-9	-15	135	81	225
;X = 740	;Y = 750	;x = 0	;y = 0	;xy = 1200	;x^2 = 1290	;y^2 = 1450
$\bar{X}$ = 74.0	$\bar{Y}$ = 75.0					

$$K_{xy} = .88$$

$$K^2_{xy} = .77$$ → AMOUNT OF VARIATION IN "X" THAT CAN BE EXPLAINED BY "Y"

Regression Analysis

1. BIVARIATE REGRESSION - MULTIPLE REGRESSION

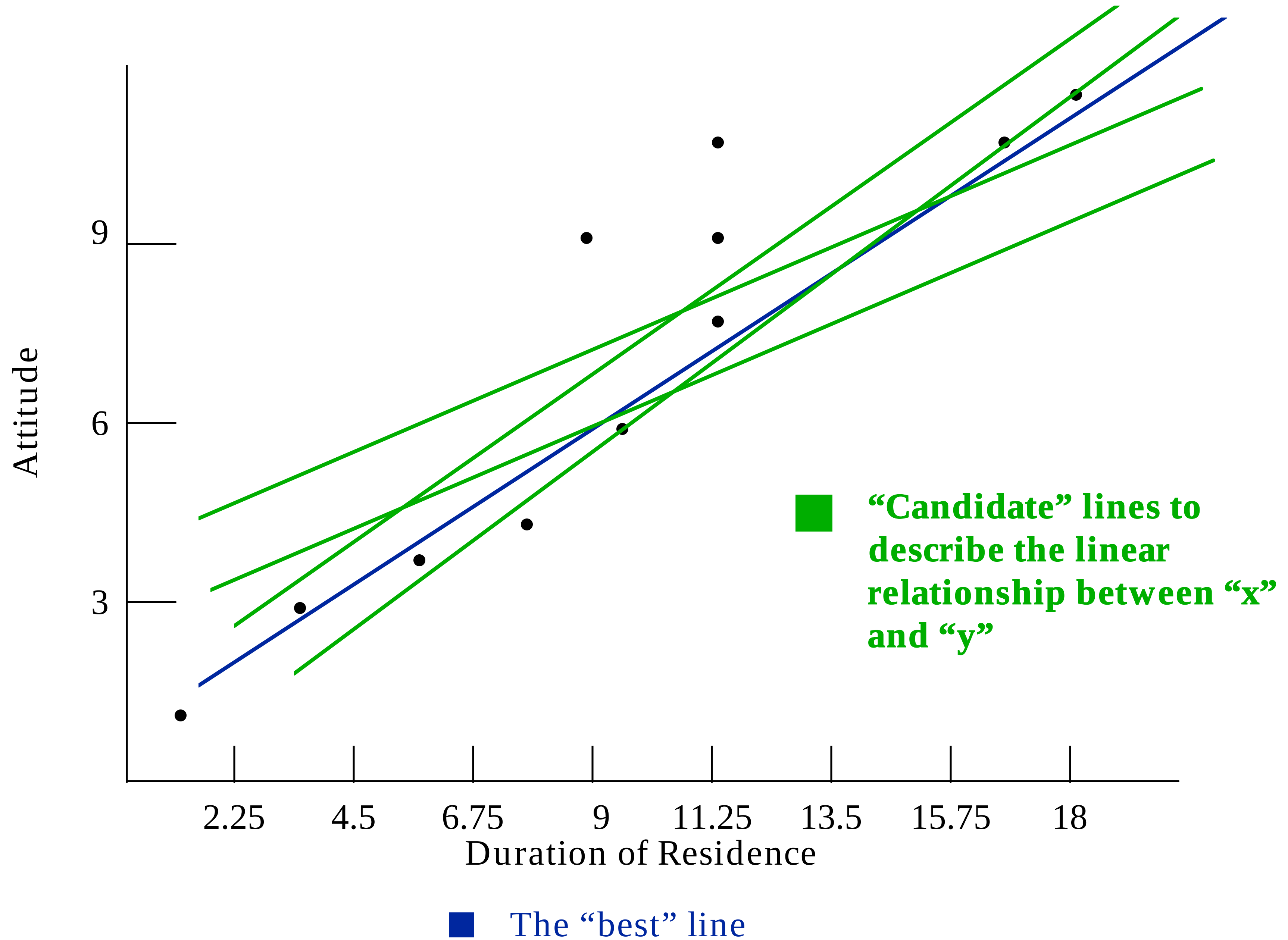
   1D, 1I   1D, nI

2. PURPOSE OF REGRESSION ANALYSIS?

   - Describe relationship(s) between (among) independent variable(s) and dependent variable
   - Predict dependent variable from a knowledge of the independent variables

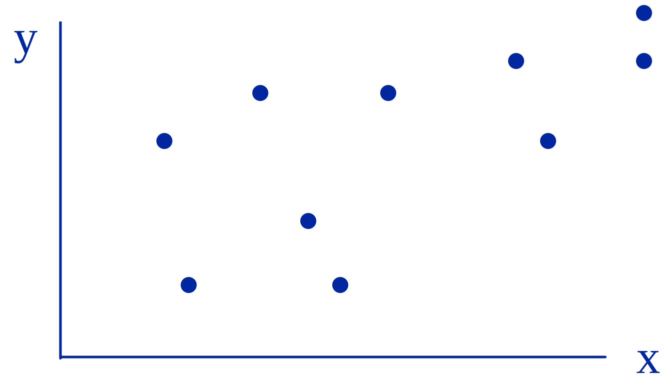
3. PURPOSE IS ACHIEVED BY "FITTING" LINEAR RELATIONSHIPS BETWEEN (AMONG) THE VARIABLES OF INTEREST

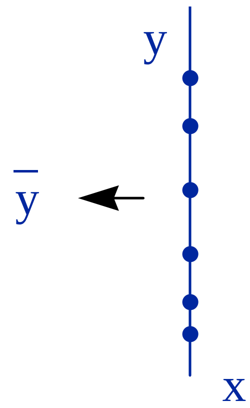
FIGURE 19.3 PLOT OF ATTITUDE WITH DURATION



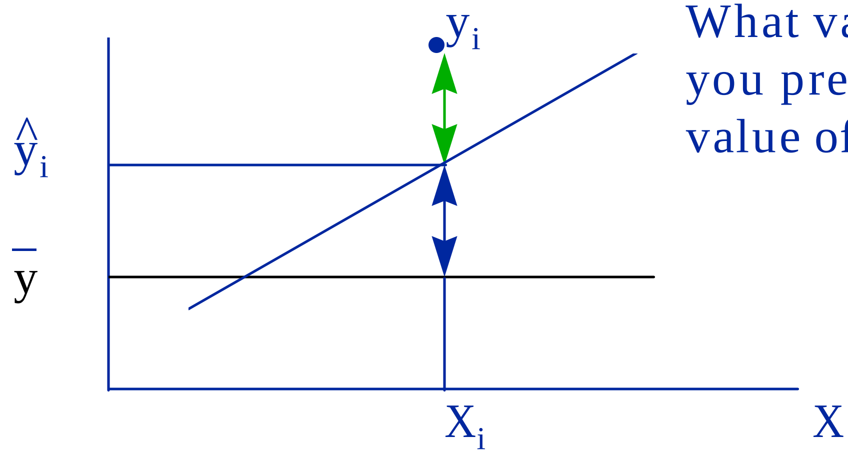
Main issue in regression analysis: How to Define "Best"

THE "ESSENCE" OF REGRESSION ANALYSIS





What value of "y" would you predict for a given value of "x"?

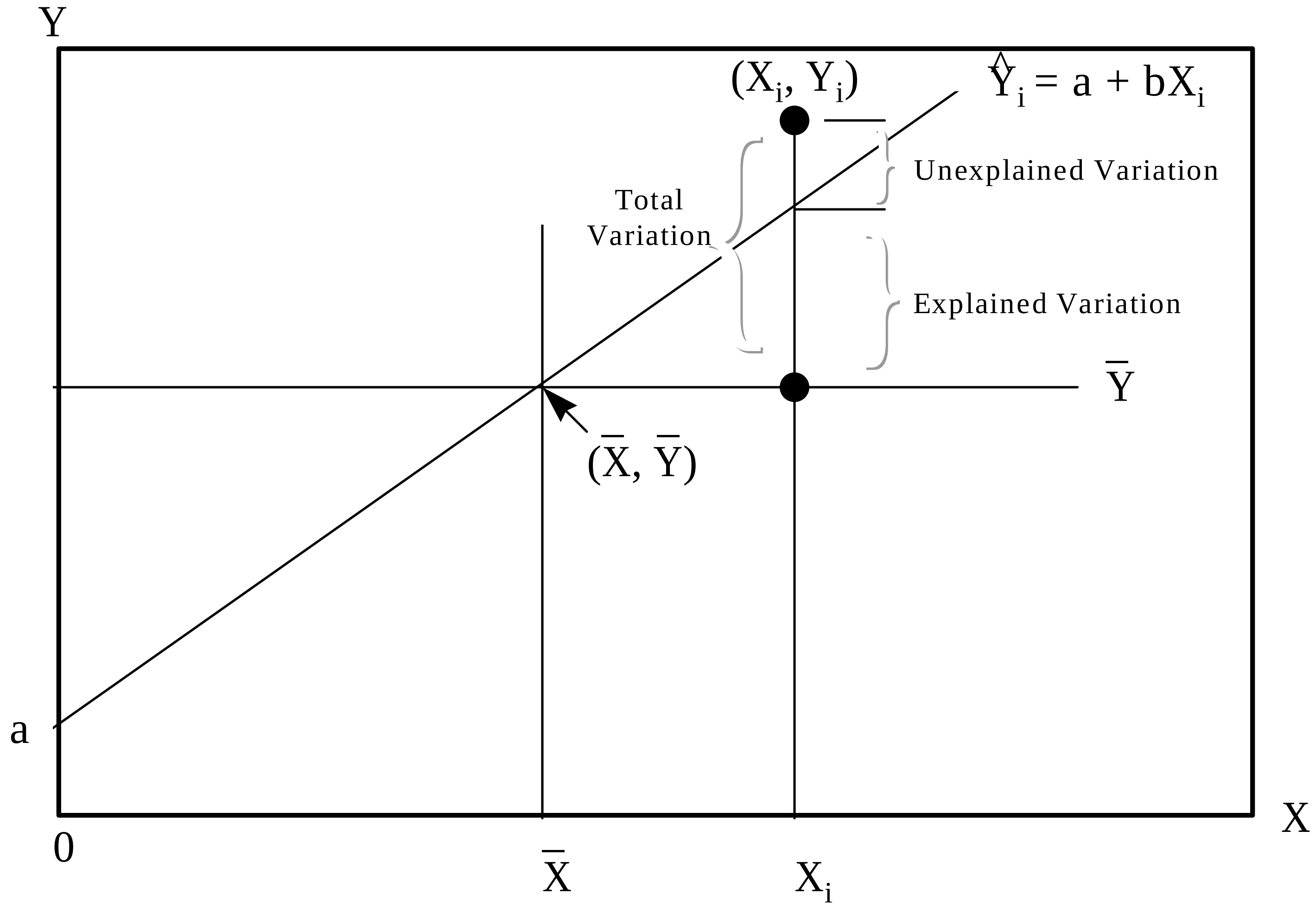


What value of "y" would you predict for a given value of "x"?

"IMPROVEMENT" IN PREDICTION = $\hat{y}_i - \bar{y}$

(BUT WE ARE STILL UNABLE TO PREDICT "Y" EXACTLY)

Measures of Variation in a Regression



$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

Total Variation	Explained Variation	Unexplained Variation

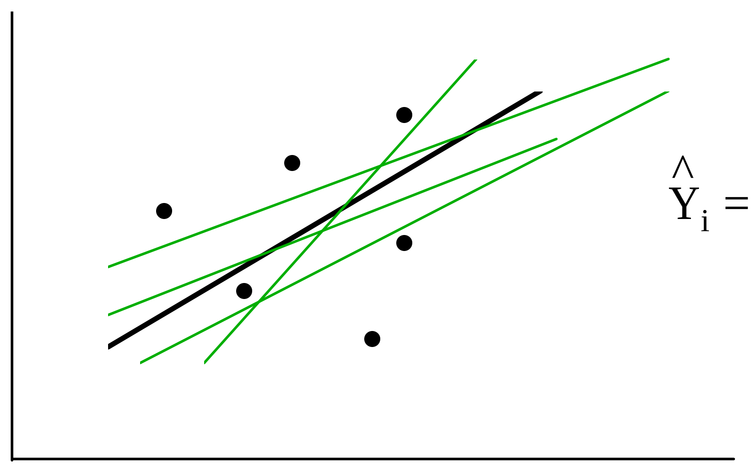
NOTE: Do not want ⊕ and ⊖ deviations to cancel each other out

TOTAL VARIATION = EXPLAINED VARIATION + UNEXPLAINED VARIATION

; $(Y_i - \bar{Y})^2$ = ; $(\hat{Y}_i - \bar{Y})^2$ + ; $(Y_i - \hat{Y}_i)^2$

(SS_{TOTAL}) (SS_{REG}) (SS_{ERROR})

Thus far we have assumed that the regression line (i.e., the $\hat{Y}$s) is known. Now we drop that assumption and go out and calculate one!



$\hat{Y}_i = a + bX_i$

(Pg. 559)

CRITERIA TO USE? MINIMIZE UNEXPLAINED VARIATION OR SS_{ERROR}

(Pg. 560) The regression line is known the moment "a" and "b" are known

Coefficient of Determination (R^2)

The percent of the total variation in the dependent variable explained by the independent variable.

$$R^2 = \frac{\text{EXPLAINED VARIATION}}{\text{TOTAL VARIATION}}$$

$$= \frac{SS_{REG}}{SS_{TOTAL}}$$

Recall, criteria for selecting the "best" line (i.e., the regression line) is to minimize unexplained variation

A SAMPLE REGRESSION CALCULATION

TABLE 19.1 EXPLAINING ATTITUDE TOWARD THE CITY OF RESIDENCE

Respondent No.	Attitude Toward the City (Y)	Duration of Residence (X)	Importance Attached to Weather
1	6	10	3
2	9	12	11
3	8	12	4
4	3	4	1
5	10	12	11
6	4	6	1
7	5	8	7
8	2	2	4
9	11	18	8
10	9	9	10
11	10	17	8
n = 12	2	2	5

$$b = \sum_{i=1}^{n} \frac{(X_i - \bar{X})\ (Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = .5897$$

$$a = \bar{Y} - b\bar{X} = 1.0793$$

NOTE: “a” and “b” satisfy criteria mentioned earlier

REGRESSION EQUATION:
ATTITUDE ($\bar{Y}$) = 1.0793 + 0.5897 (DURATION OF RESIDENCE)

TABLE 19.2 BIVARIATE REGRESSION

Multiple R	.93608
R^2	.87624
Adjusted R^2	.86387
Standard Error	1.22329

$R^2 = \frac{\text{EXPLAINED VARIATION}}{\text{TOTAL VARIATION}}$

Analysis of Variance

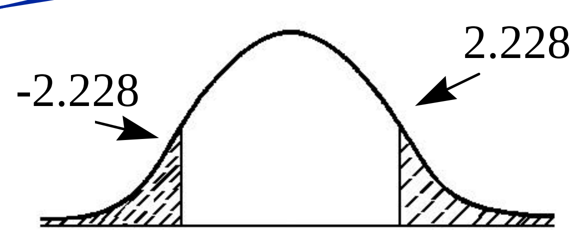
		Sum of Squares	Mean Square
Regression	df	105.95222 (explained)	105.95222
Residual (ERROR)	1	14.96444 (unexplained)	1.49644
F = 70.80266	10	120.9168 (total)	

Signif *F* - .0000

VARIABLES IN THE EQUATION

Variable	*b*	SE*b*	Beta (*B*)	*T*	Sig. *T*
Duration	.58972	.07008	.93608	8.414	.0000
(Constant)	1.07932	.74335		1.452	.1772

$$t = \frac{b}{SE_b}$$



(TABLE 4)