Chapter 14
Simple Linear Regression

Outline

- ✓The Simple Linear Regression Model
  - ✓Regression Model the Regression Equations
  - ✓Estimated Regression Equation
- ✓The Least Squares Method
- ✓The Coefficient of Determination
  - ✓Correlation coefficient

Measures of Association Between Two Variables

- ✓ Chapter 3 section 5
- ✓ Covariance
- ✓ Correlation Coefficient

Covariance – conceptual formula

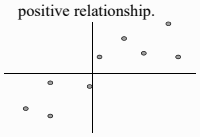
Sample Covariance

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

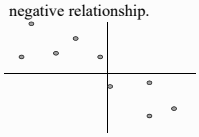
Sample variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Positive values indicate a positive relationship.



Negative values indicate a negative relationship.

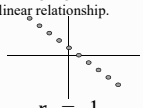


Correlation Coefficient

Sample Correlation Coefficient $r_{xy} = \frac{s_{xy}}{s_x s_y}$

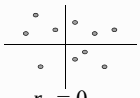
The coefficient can take on values between -1 and +1: $-1 \le r_{xy} \le 1$

Values near **-1** indicate a strong negative linear relationship.



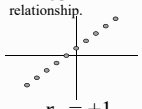
$r_{xy} = -1$

Values near **0** indicate no linear relationship.



$r_{xy} = 0$

Values near **+1** indicate a strong positive linear relationship.



$r_{xy} = +1$

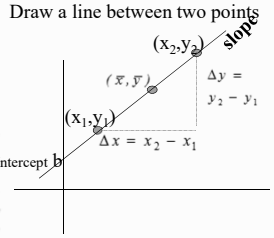
Equations for calculation

Sample covariance applied equation

$$s_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n(n-1)}$$

Algebra 1 Review

Draw a line between two points



General line equation

$y = mx + b$

where,

m is the slope

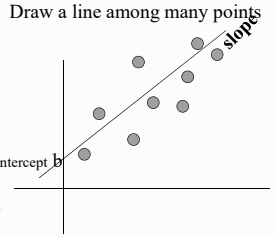
$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

b is the intercept
(you need the slope and one point)

$$b = \bar{y} - m\bar{x}$$

Simple Linear Regression

Draw a line among many points



General line equation

$\hat{y}_i = b_0 + b_1 x_i$

where,

b_1 is the estimated slope

$$b_1 = \frac{\text{cov}(x,y)}{\text{var}(x)} = \frac{s_{xy}}{s_x^2}$$

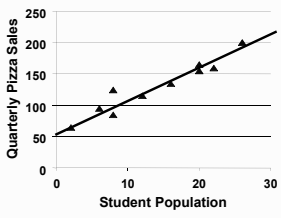
b_0 is the intercept
(you need the slope and one point)

$$b_0 = \bar{y} - b_1\bar{x}$$

Example: Armand's Pizza

x_i Student Population (000)	y_i Quarterly Sales ($000)
2	58
6	105
8	88
8	118
12	117
16	137
20	157
20	169
22	149
26	202

$\bar{x} = 14$ $\bar{y} = 130$



Example: Armand's Pizza

x Pop (000)	y Sales ($000)	x^2	y^2	xy
2	58	4	3364	116
6	105	36	11025	630
8	88	64	7744	704
8	118	64	13924	944
12	117	144	13689	1404
16	137	256	18769	2192
20	157	400	24649	3140
20	169	400	28561	3380
22	149	484	22201	3278
26	202	676	40804	5252
140	**1300**	**2528**	**184730**	**21040**

Determine the slope

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{315.5}{63.1} = 5$$

$$s_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{n-1} = 315.5$$

$$s_x^2 = 63.1$$

Example: Armand's Pizza

x Pop (000)	y Sales ($000)	x^2	y^2	xy
2	58	4	3364	116
6	105	36	11025	630
8	88	64	7744	704
8	118	64	13924	944
12	117	144	13689	1404
16	137	256	18769	2192
20	157	400	24649	3140
20	169	400	28561	3380
22	149	484	22201	3278
26	202	676	40804	5252
140	**1300**	**2528**	**184730**	**21040**

$\bar{x} = 140 / 10 = 14$

$\bar{y} = 1300 / 10 = 130$

Determine the intercept

$b_0 = \bar{y} - b_1 \cdot \bar{x}$

$b_0 = 130 - 5 \cdot 14 = \mathbf{60}$

Write the regression equation

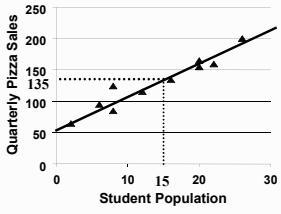
$\hat{y}_i = b_0 + b_1 x_i$

$$\hat{y}_i = 60 + 5 x_i$$

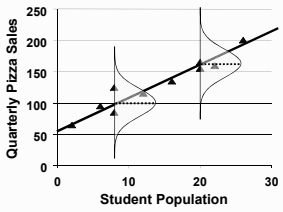
$\hat{y}_i = 60 + 5 x_i$

What is the estimated quarterly pizza sales for a store with 8 thousand students on campus?

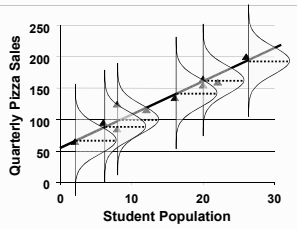
with 15 thousand students?



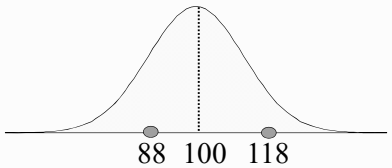
$\hat{y}_i = 60 + 5x_i$



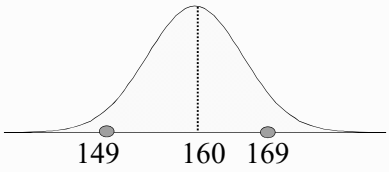
$\hat{y}_i = 60 + 5x_i$



8 thousand students



20 thousand students



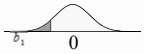
Seven Steps of Hypothesis Testing

Where: $\beta_1 > 0$ You believe there is a positive relationship between X and Y

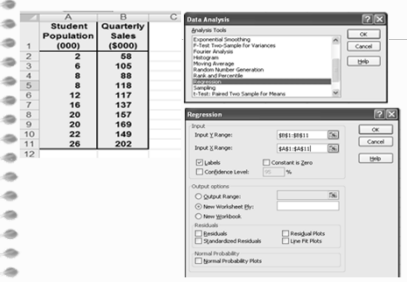
1. Ho : $\beta_1 \leq 0$
   Ha : $\beta_1 > 0$
4. 

   Reject Ho if P - value $< \alpha$
5. Use Excel

Where: $\beta_1 < 0$ You believe there is a negative relationship between X and Y

1. Ho : $\beta_1 \geq 0$
   Ha : $\beta_1 < 0$
4. 

   Reject Ho if P - value $< \alpha$
5. Use Excel



	A	B
1	Student Population (000)	Quarterly Sales ($000)
2	2	58
3	6	105
4	8	88
5	8	118
6	12	117
7	16	137
8	20	157
9	20	169
10	22	149
11	26	202
12		

EXCEL Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95012296
R Square	0.90273363
Adjusted R Square	0.89057533
Standard Error	13.8293167
Observations	10

ANOVA

	df	SS	MS	F	Significance F
Regression	1	14200	14200	74.248366	2.54887E-05
Residual	8	1530	191.25		
Total	9	15730			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	60	9.22603481	6.503336	0.0001874	38.72473	81.27527
Student Population (000)	5	0.580265238	8.616749	2.549E-05	3.661906	6.338094

6. .0000255 < .01,Reject Ho, Accept Ha
   The sign of the slope must correspond to the alternate hypothesis
7. Student population is significant.