The Regression Line
Sections 5.1, 5.2
Lecture 13

Robb T. Koether

Hampden-Sydney College

Thu, Feb 4, 2016

Outline

1. The Regression Line
2. Review of Straight Lines
3. The Least-Squares Regression Line
4. Example
5. Assignment

Outline

1. The Regression Line
2. Review of Straight Lines
3. The Least-Squares Regression Line
4. Example
5. Assignment

- The regression line is the line of "best fit."
- We will see soon what "best fit" means.
- In the example of the free-lunch rate vs. the graduation rate, the regression line turns out to be

$$y = 91.047 - 0.494x$$

where x is the free-lunch rate and y is the graduation rate.

- The regression line is the line of “best fit.”
- We will see soon what “best fit” means.
- In the example of the free-lunch rate vs. the graduation rate, the regression line turns out to be

$$y = 91.047 - 0.494x$$

where x is the free-lunch rate and y is the graduation rate.
- What does that tell us?

Outline

1. The Regression Line
2. Review of Straight Lines
3. The Least-Squares Regression Line
4. Example
5. Assignment

The Equation of a Line

- The equation of a (nonvertical) straight line is of the form

$$y = a + bx.$$

- The coefficient a is the y-intercept.
- The coefficient b is the slope.
- The variable x is the explanatory variable.
- The variable y is the response variable.

Slope and Intercept

- The y-intercept is the value of y when $x = 0$.
- Normally, but not always, this is well out of the range of interest.
- The slope tells us how fast y changes as x changes.
- Specifically, a slope of b means that if x increases by 1, then y will change by b.

Example

Example

District	Free Lunch	Grad. Rate	District	Free Lunch	Grad. Rate
Amelia	41.2	68.9	King and Queen	59.9	64.1
Caroline	40.2	62.9	King William	27.9	67.0
Charles City	45.8	67.7	Louisa	44.9	80.1
Chesterfield	22.5	80.5	New Kent	13.9	77.0
Colonial Hgts	25.7	73.0	Petersburg	61.6	54.6
Cumberland	55.3	63.9	Powhatan	12.2	89.3
Dinwiddie	45.2	71.4	Prince George	30.9	85.0
Goochland	23.3	76.3	Richmond	74.0	46.9
Hanover	13.7	90.1	Sussex	74.8	59.0
Henrico	30.2	81.1	West Point	19.1	82.0
Hopewell	63.1	63.4			

Example

Example (Slope and Intercept)

- Use the equation
$$y = 91.047 - 0.494x$$
to predict the graduation rate when the free-lunch rate is
  - 20%
  - 50%
  - 100%
  - 44.9% (Louisa)

Example

Example (Slope and Intercept)

- Use the equation

$$y = 91.047 - 0.494x$$

  to predict the graduation rate when the free-lunch rate is
  - 20%
  - 50%
  - 100%
  - 44.9% (Louisa)
- What does the y-intercept tell us?

Example

Example (Slope and Intercept)

- Use the equation

$$y = 91.047 - 0.494x$$

  to predict the graduation rate when the free-lunch rate is
  - 20%
  - 50%
  - 100%
  - 44.9% (Louisa)
- What does the y-intercept tell us?
- What does the slope tell us?

Outline

1 The Regression Line

2 Review of Straight Lines

3 The Least-Squares Regression Line

4 Example

5 Assignment

The Predicted Response

- We use the variable $\hat{y}$, read "y hat," to indicate the predicted y value.
- The variable y is the observed y value.
- The equation

$$\hat{y} = a + bx$$

allows us to find the predicted value of y for a given value of x.

The Error

- It is unlikely that the predicted $\hat{y}$ will agree with the observed y.
- The difference is called the error (or the residual).

$$\begin{aligned} \text{error} &= \text{observed response} - \text{predicted response} \\ &= y - \hat{y}. \end{aligned}$$

The Least-Squares Regression Line

Definition (The Least-Squares Regression Line)

The least-squares regression line is the line that makes the sum of the squared errors as small as possible.

Outline

1. The Regression Line
2. Review of Straight Lines
3. The Least-Squares Regression Line
4. Example
5. Assignment

Example

Example (Sum of Squared Errors)

- Consider the data

x	y
3	40
5	80
7	160
9	180
16	240

- The regression line turns out to be

$$y = 20 + 15x.$$

- Calculate the errors and the sum of the squared errors.
- Draw the scatterplot and indicate the errors.

Example

Example (Sum of Squared Errors)

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	40			
5	80			
7	160			
9	180			
16	240			

Example

Example (Sum of Squared Errors)

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	40	65		
5	80	95		
7	160	125		
9	180	155		
16	240	260		

Example

Example (Sum of Squared Errors)

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	40	65	-25	
5	80	95	-15	
7	160	125	35	
9	180	155	25	
16	240	260	-20	

Example

Example (Sum of Squared Errors)

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	40	65	-25	625
5	80	95	-15	225
7	160	125	35	1225
9	180	155	25	625
16	240	260	-20	400

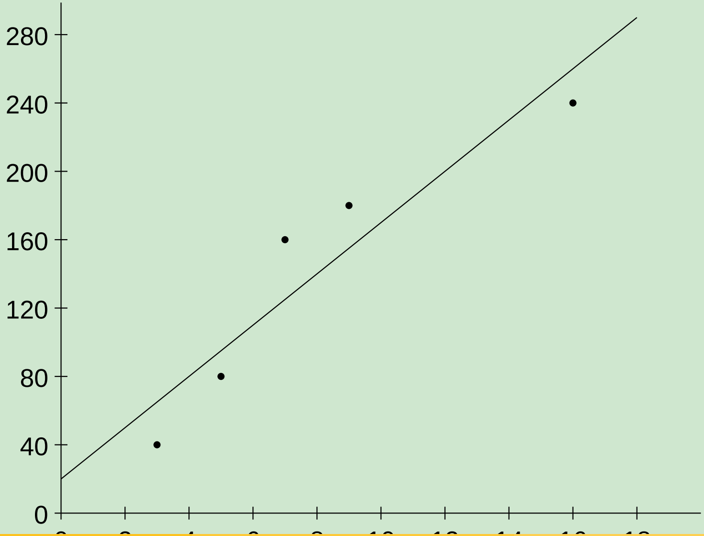
Example

Example (Sum of Squared Errors)

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	40	65	-25	625
5	80	95	-15	225
7	160	125	35	1225
9	180	155	25	625
16	240	260	-20	400
			0	3100

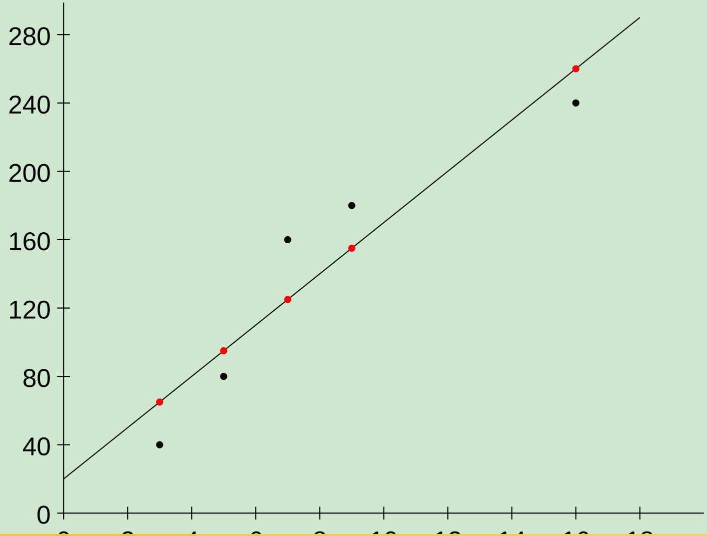
Example

Example (Sum of Squared Errors)



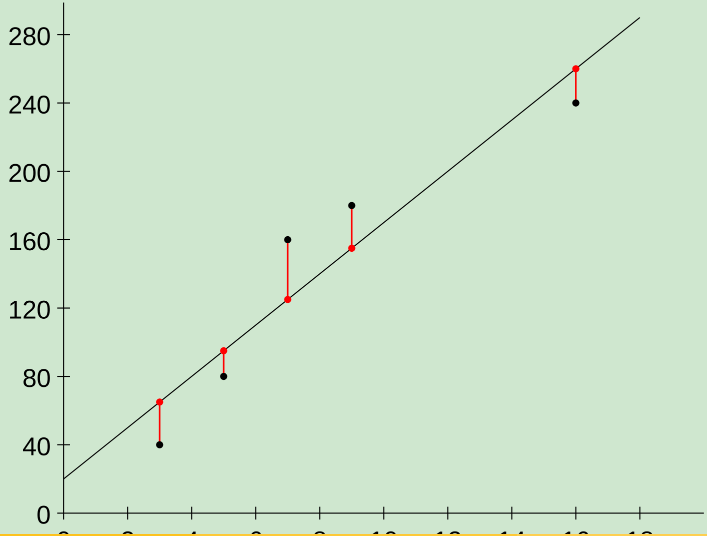
Example

Example (Sum of Squared Errors)



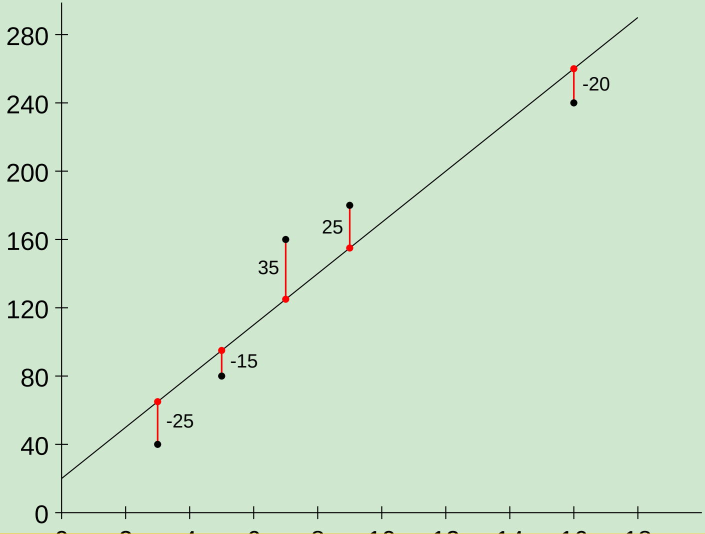
Example

Example (Sum of Squared Errors)



Example

Example (Sum of Squared Errors)



Outline

1. The Regression Line
2. Review of Straight Lines
3. The Least-Squares Regression Line
4. Example
5. Assignment

Assignment

Assignment

- Read Sections 5.1, 5.2, 5.3.
- Apply Your Knowledge: 1, 2, 4, 5.
- Check Your Skills: 20, 21, 23, 24, 25, 26.
- Exercises: 30, 34, 35(a-b), 37, 38.