Scatterplots
Sections 4.1, 4.2, 4.3
Lecture 11

Robb T. Koether

Hampden-Sydney College

Mon, Feb 1, 2016

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Bivariate Data

Definition (Bivariate)

Data are called bivariate if two observations are made for each member of the sample. We usually call the observations x and y, .

- x is the explanatory variable.
- y is the response variable.
- x is also called the independent variable.
- y is also called the dependent variable.

Explanatory and Response Variables

Definition (Response Variable)

A response variable measures an outcome of a study.

Definition (Explanatory Variable)

An explanatory variable may explain or influence the value of the response variable.

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Is the free-lunch rate in a school district correlated with the graduation rate in that district?
- The Richmond Times-Dispatch published data for school districts in the Richmond area.

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

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			

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Which is the response variable and which is the explanatory variable?
  - Free lunch rate is response, graduation rate is explanatory?

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Which is the response variable and which is the explanatory variable?
  - Free lunch rate is response, graduation rate is explanatory?
  - Graduation rate is response, free lunch rate is explanatory?

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Which is the response variable and which is the explanatory variable?
  - Free lunch rate is response, graduation rate is explanatory?
  - Graduation rate is response, free lunch rate is explanatory?
  - Both are response variables?

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Which is the response variable and which is the explanatory variable?
  - Free lunch rate is response, graduation rate is explanatory?
  - Graduation rate is response, free lunch rate is explanatory?
  - Both are response variables?
  - Both are explanatory variables?

Free Lunches vs. Graduation Rates

Example (Free-lunch Rate vs. Graduation Rate)

- Which is the response variable and which is the explanatory variable?
  - Free lunch rate is response, graduation rate is explanatory?
  - Graduation rate is response, free lunch rate is explanatory?
  - Both are response variables?
  - Both are explanatory variables?
  - None of the above?

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Definition (Scatterplot)

A scatterplot is a graphical display in which

- The values of the explanatory variable appear on the x-axis.
- The values of the response variable appear on the y-axis.
- The observations (x-y pairs) are plotted as points on the graph.

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Interpreting Scatterplots

- We look for a pattern in the scatterplot.
- In particular, we look for three things:
  - The direction of the pattern (e.g., rising, falling).
  - The form of the pattern (e.g., linear, oval, circular).
  - The strength of the pattern (e.g., strong, weak).
- Begin by drawing (or imagining) an oval surrounding the points, excluding outliers.

Direction

- If the oval is tilted upward from left to right, then the association is positive.
- If it tilts downward, the association is negative.

Form

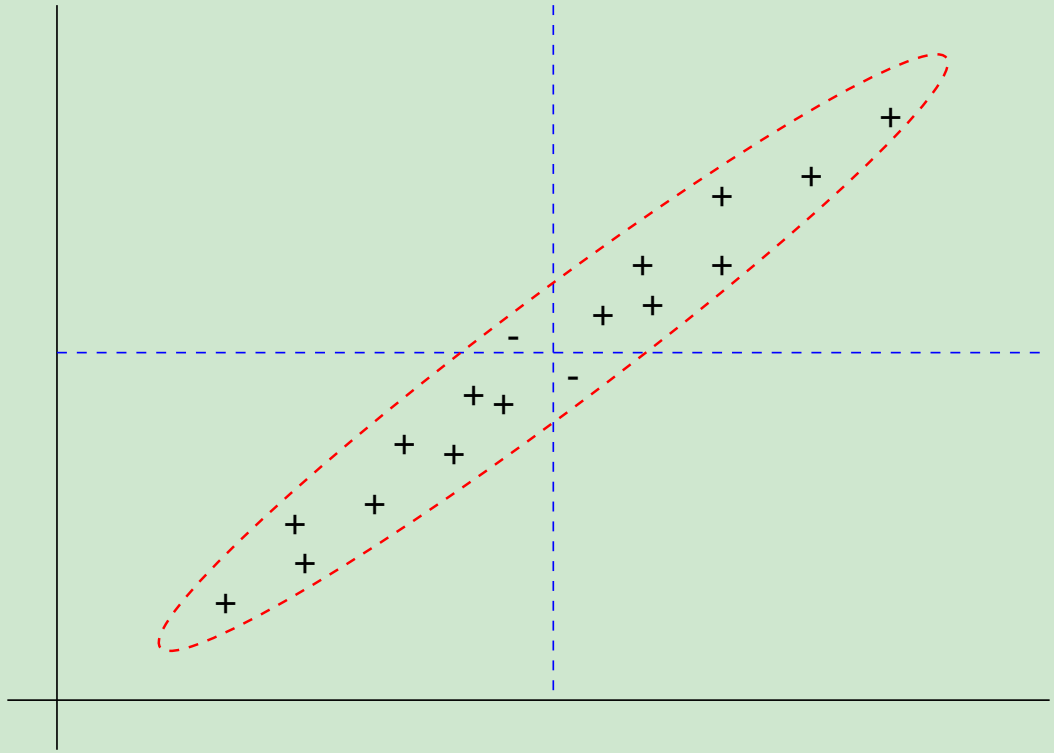
- The association is linear if the points show no tendency to curve left or right from their general direction.
- If the points do show a tendency to curve, then the association is non-linear. It may be quadratic or some other non-linear shape.

Strong vs. Weak Association

- If the oval is narrow, then the association is strong.
- If the oval is wide, then the association is weak.
- If the oval is really a circle, then there is no association.

Interpreting Scatterplots

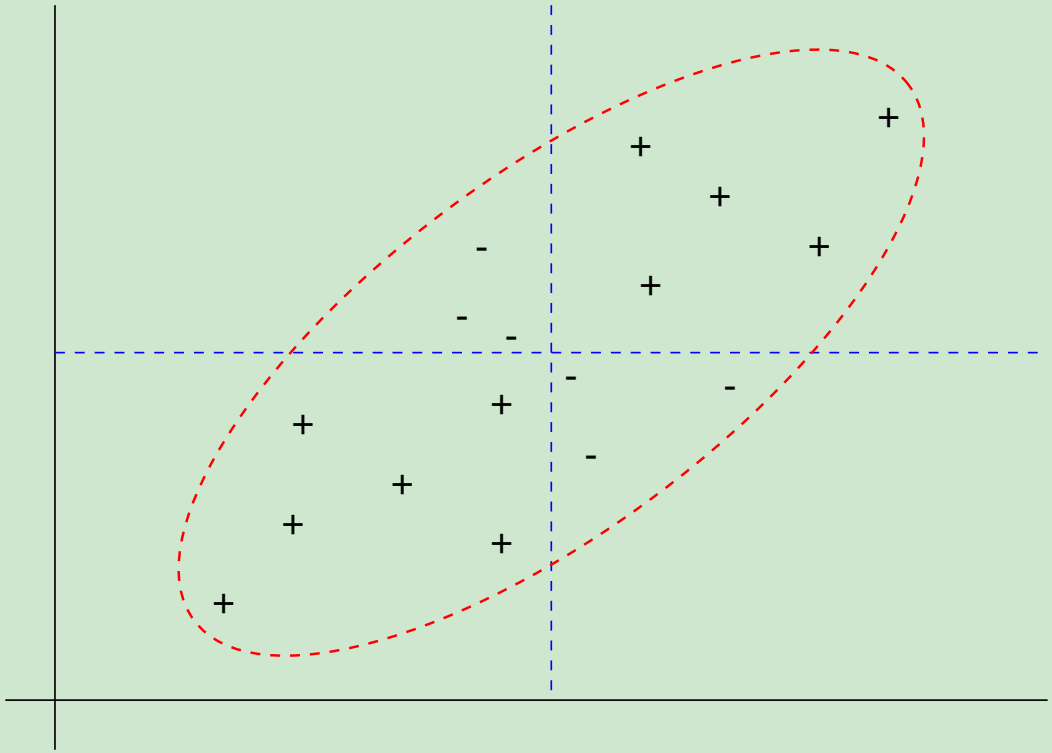
Example (Interpreting Scatterplots)



Strong positive association

Interpreting Scatterplots

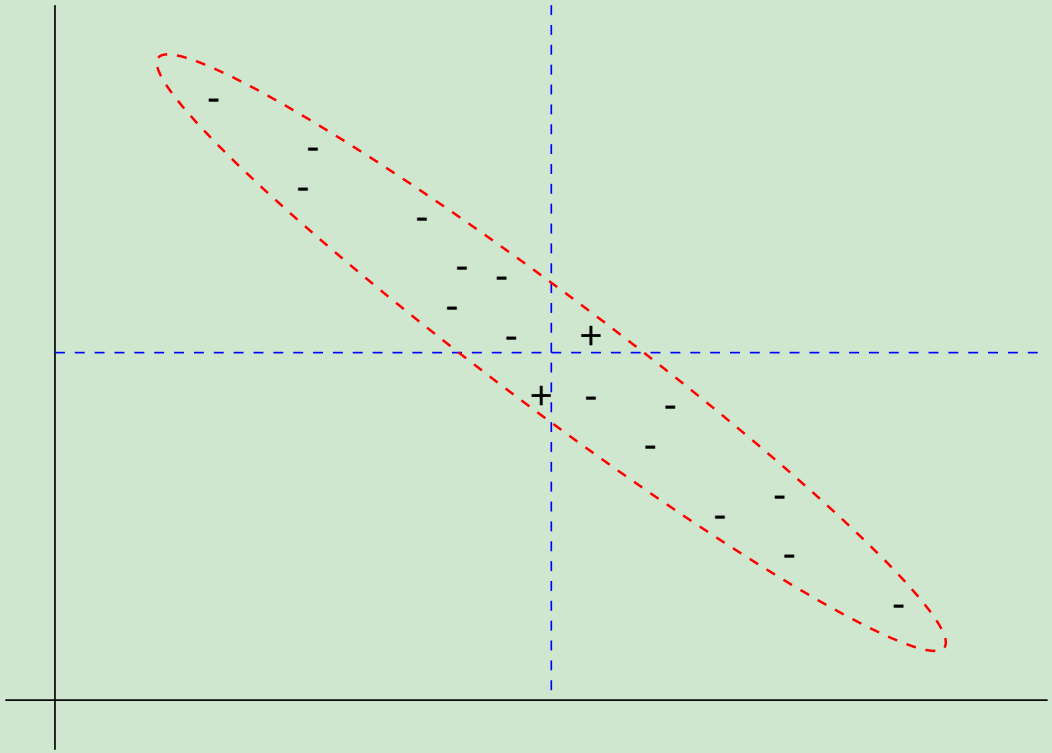
Example (Interpreting Scatterplots)



Weak positive association

Interpreting Scatterplots

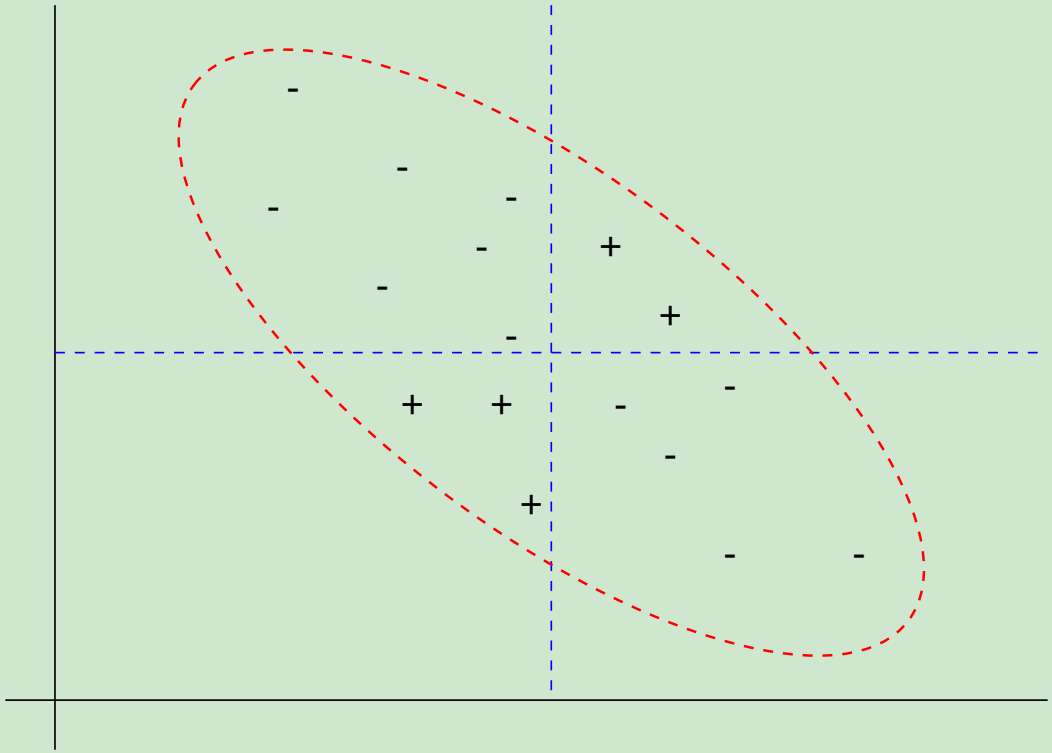
Example (Interpreting Scatterplots)



Strong negative association

Interpreting Scatterplots

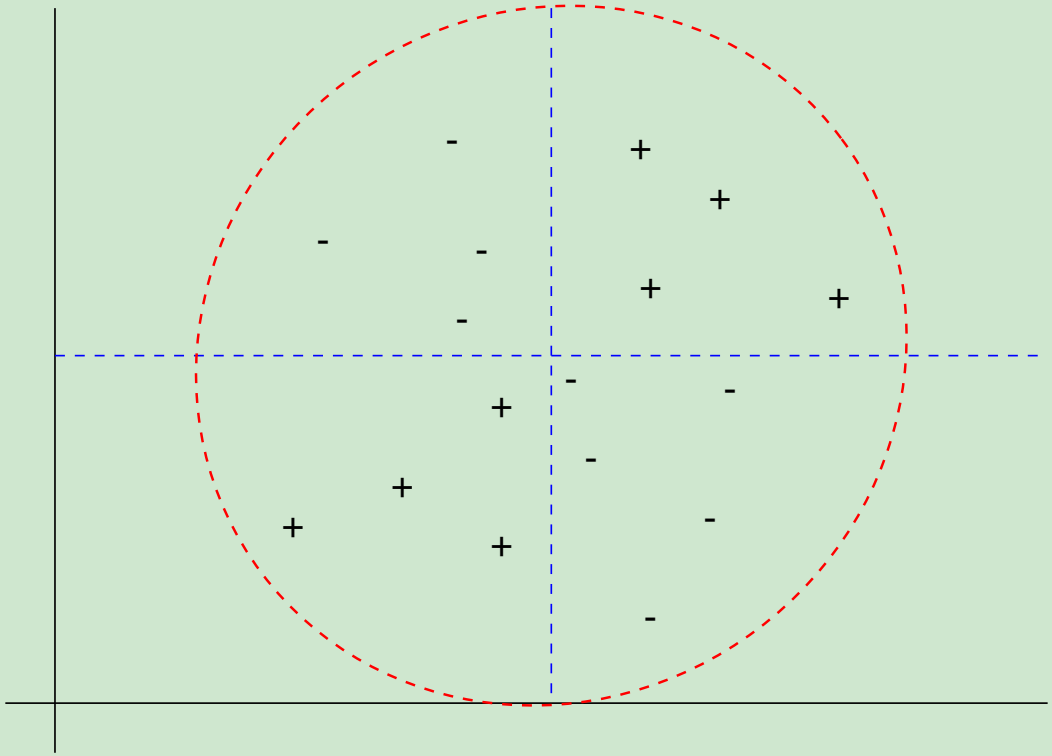
Example (Interpreting Scatterplots)



Weak negative association

Interpreting Scatterplots

Example (Interpreting Scatterplots)



No association

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Correlation

Definition (Correlation)

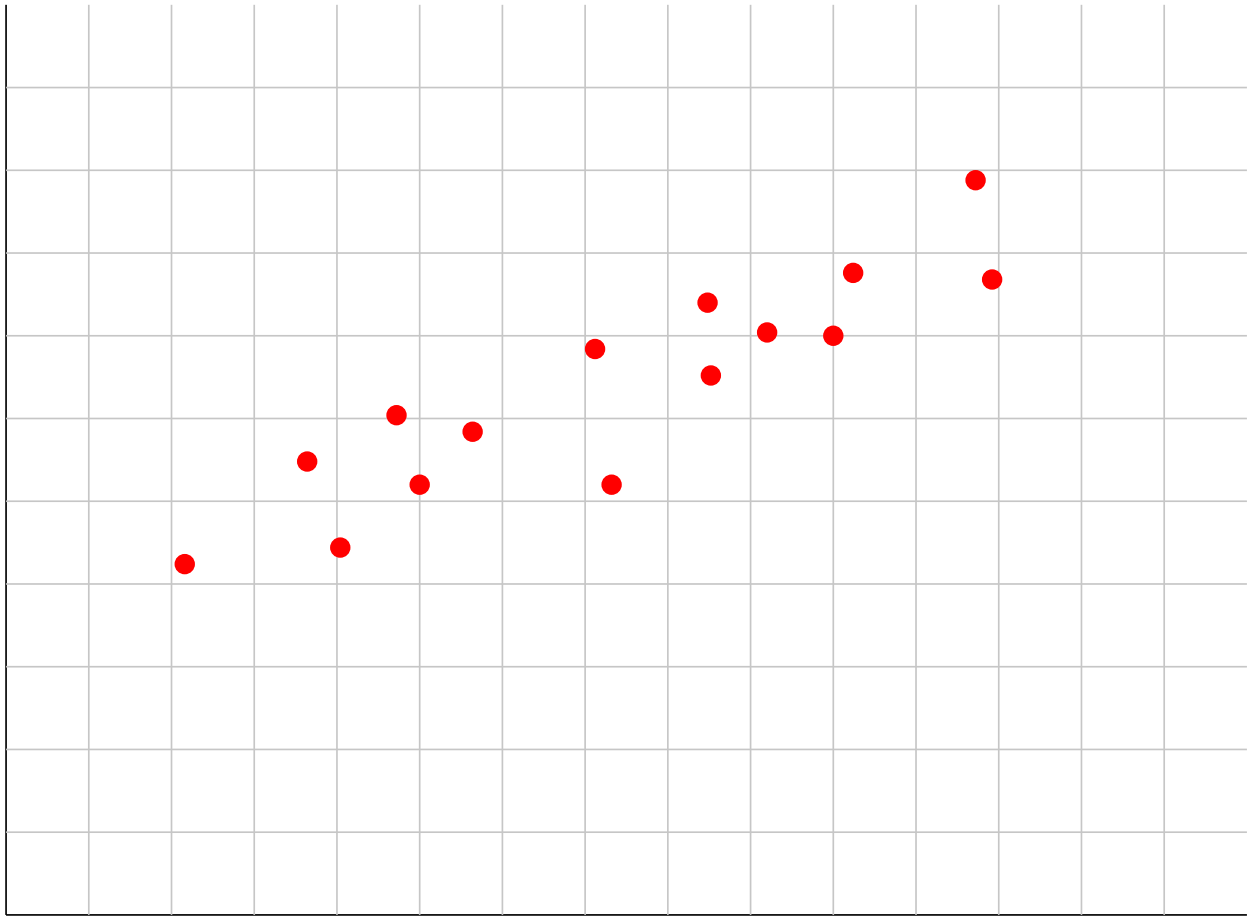
The correlation, denoted by r, of bivariate data measures the strength and the direction of the relationship (assuming that it is linear). The correlation is computed as

$$r = \frac{\sum \left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right)}{n-1}.$$

Correlation

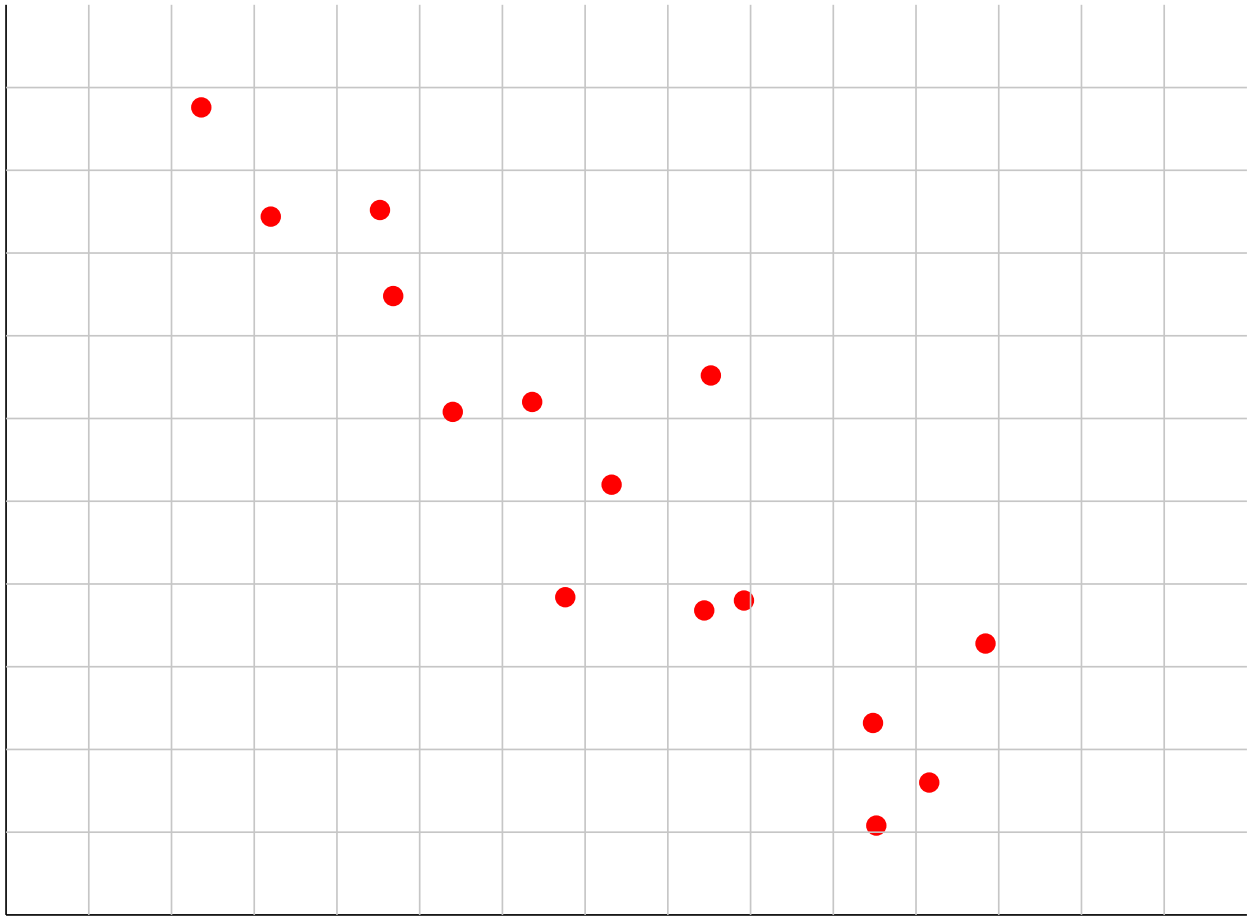
- If r is close to $+1$ or -1, the relationship is strong.
- If r is close to 0, the relationship is weak.
- In other words, $|r|$ measures the strength of the relationship.

Strong Positive Linear Association



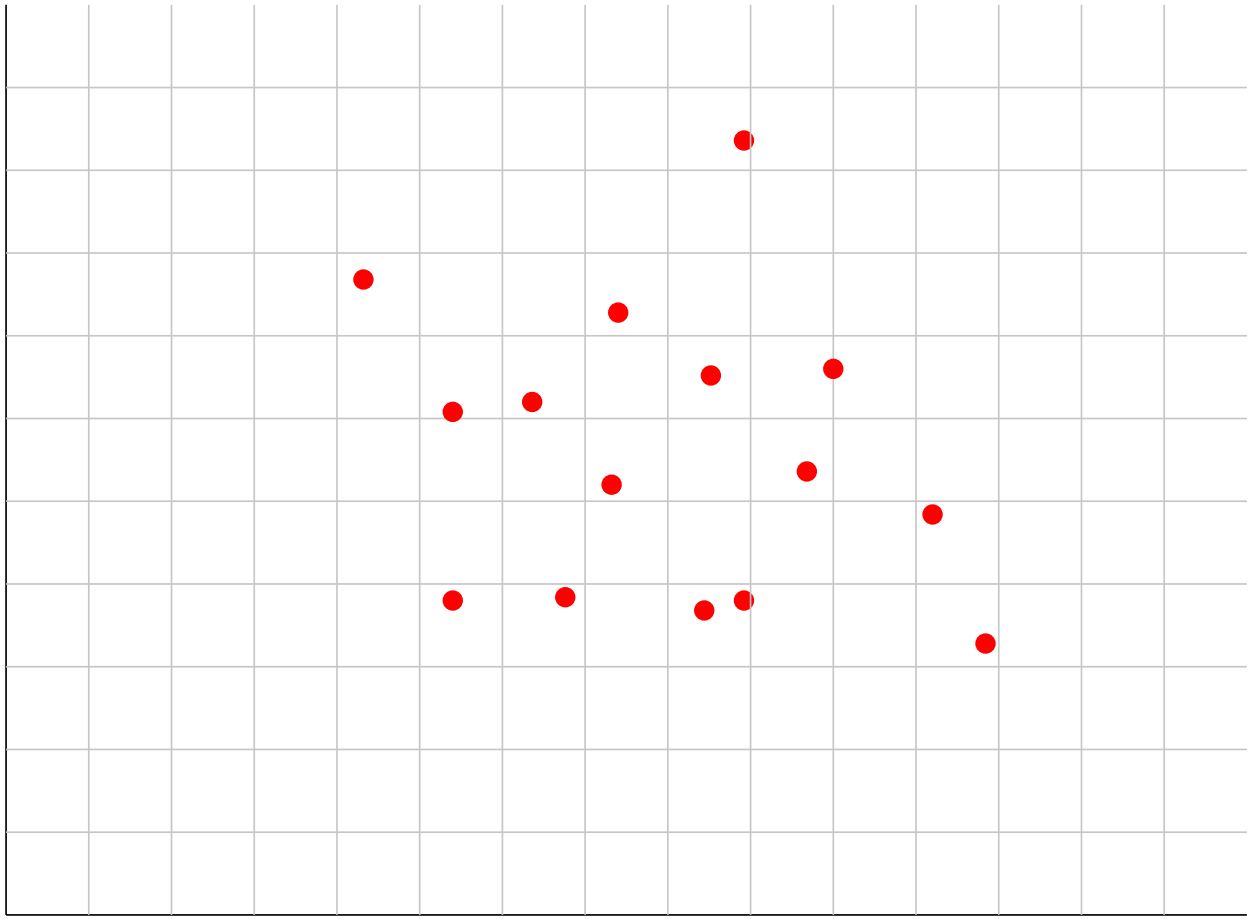
r is close to $+1$

Strong Positive Linear Association



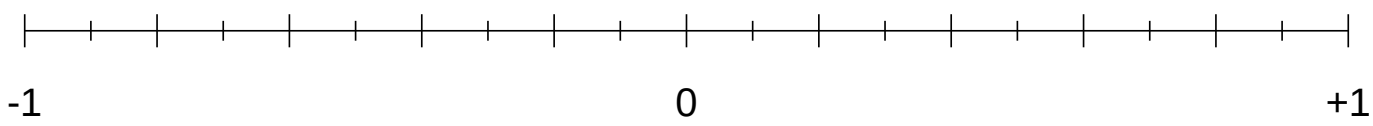
r is close to -1

Strong Positive Linear Association

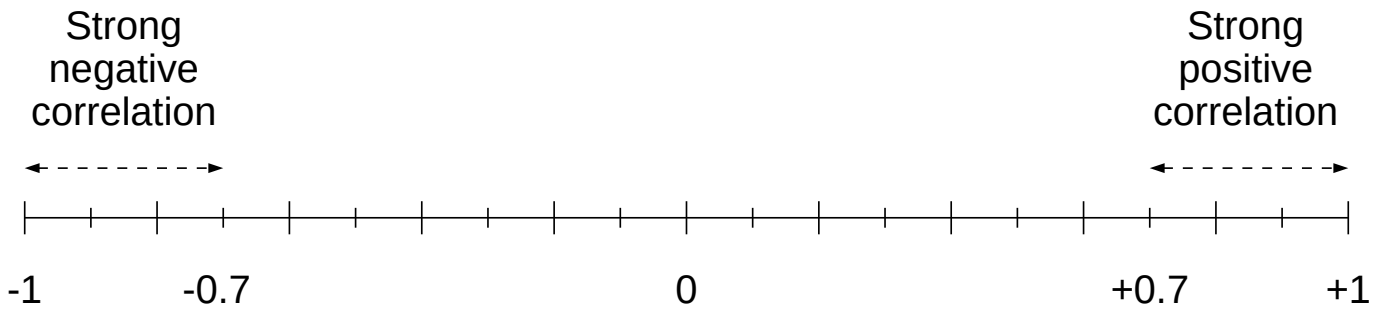


r is close to 0

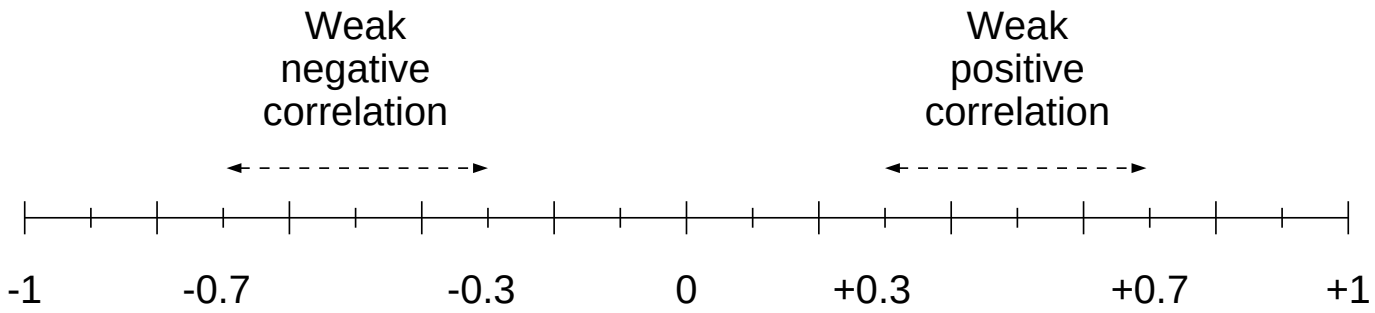
Strong Positive Linear Association



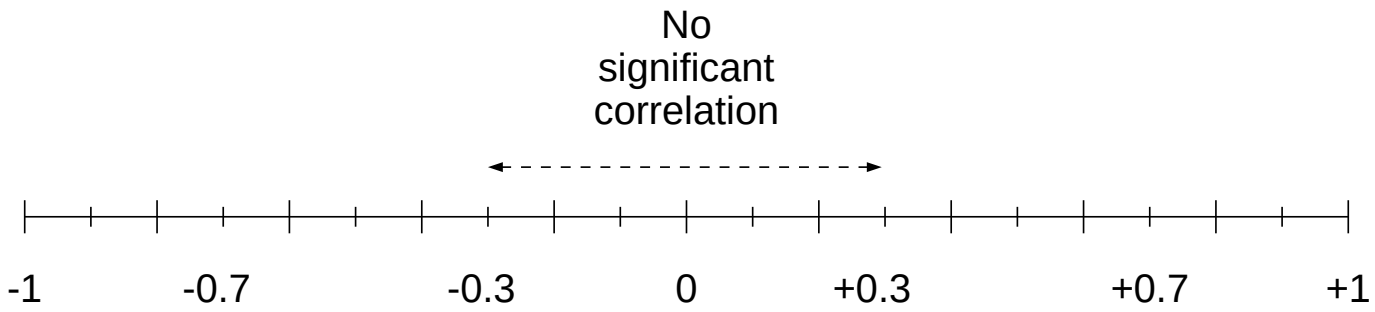
Strong Positive Linear Association



Strong Positive Linear Association



Strong Positive Linear Association



Example

Example (Height vs. Weight)

- Consider the following height and weight data.

Height (x)	Weight (y)
70	185
65	140
71	180
76	220
68	150
67	170
68	185
72	205
74	210
69	155

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185					
65	140					
71	180					
76	220					
68	150					
67	170					
68	185					
72	205					
74	210					
69	155					

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0				
65	140	−5				
71	180	1				
76	220	6				
68	150	−2				
67	170	−3				
68	185	−2				
72	205	2				
74	210	4				
69	155	−1				

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0	0.0			
65	140	−5	−1.5			
71	180	1	0.3			
76	220	6	1.8			
68	150	−2	−0.6			
67	170	−3	−0.9			
68	185	−2	−0.6			
72	205	2	0.6			
74	210	4	1.2			
69	155	−1	−0.3			

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0	0.0	5		
65	140	−5	−1.5	−40		
71	180	1	0.3	0		
76	220	6	1.8	40		
68	150	−2	−0.6	−30		
67	170	−3	−0.9	−10		
68	185	−2	−0.6	5		
72	205	2	0.6	25		
74	210	4	1.2	30		
69	155	−1	−0.3	−25		

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0	0.0	5	0.1875	
65	140	−5	−1.5	−40	−1.5000	
71	180	1	0.3	0	0.0000	
76	220	6	1.8	40	1.5000	
68	150	−2	−0.6	−30	−1.1250	
67	170	−3	−0.9	−10	−0.3750	
68	185	−2	−0.6	5	0.1875	
72	205	2	0.6	25	0.9375	
74	210	4	1.2	30	1.1250	
69	155	−1	−0.3	−25	−0.9375	

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0	0.0	5	0.1875	0.0000
65	140	−5	−1.5	−40	−1.5000	2.2500
71	180	1	0.3	0	0.0000	0.0000
76	220	6	1.8	40	1.5000	2.7000
68	150	−2	−0.6	−30	−1.1250	0.6750
67	170	−3	−0.9	−10	−0.3750	0.3375
68	185	−2	−0.6	5	0.1875	−0.1125
72	205	2	0.6	25	0.9375	0.5625
74	210	4	1.2	30	1.1250	1.3500
69	155	−1	−0.3	−25	−0.9375	0.28125

Example

Example (Height vs. Weight)

- Compute the deviations, their products, and the sums.

her

Height (x)	Weight (y)	$x-\overline{x}$	$\frac{x-\overline{x}}{s_x}$	$y-\overline{y}$	$\frac{y-\overline{y}}{s_y}$	$\left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$
70	185	0	0.0	5	0.1875	0.0000
65	140	−5	−1.5	−40	−1.5000	2.2500
71	180	1	0.3	0	0.0000	0.0000
76	220	6	1.8	40	1.5000	2.7000
68	150	−2	−0.6	−30	−1.1250	0.6750
67	170	−3	−0.9	−10	−0.3750	0.3375
68	185	−2	−0.6	5	0.1875	−0.1125
72	205	2	0.6	25	0.9375	0.5625
74	210	4	1.2	30	1.1250	1.3500
69	155	−1	−0.3	−25	−0.9375	0.28125
						8.04375

Example

Example (Height vs. Weight)

- Now compute

$$r = \frac{8.04375}{9} = 0.89375.$$

- We may simplify the formula:

$$r = \frac{\sum \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)}{n - 1}$$

- We may simplify the formula:

$$
\begin{aligned}
r &= \frac{\sum \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)}{n-1} \\
&= \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{(n-1)s_x s_y}
\end{aligned}
$$

- We may simplify the formula:

$$
\begin{aligned}
r &= \frac{\sum \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)}{n-1} \\
&= \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{(n-1) s_x s_y} \\
&= \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}}
\end{aligned}
$$

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Practice

- Find the correlation of the data

x	y
2	3
7	4
9	8

Outline

1. Explanatory and Response Variables
2. Scatterplots
3. Interpreting Scatterplots
4. Measuring Linear Association
5. Practice
6. Assignment

Assignment

Assignment

- Read Sections 4.1, 4.2, 4.3, 4.5.
- Apply Your Knowledge: 1, 3, 4, 6.
- Check Your Skills: 14, 15, 19, 20.
- Exercises: 24, 26, 27, 28, 29.