Correlation
Sections 4.5, 4.6
Lecture 12

Robb T. Koether

Hampden-Sydney College

Wed, Feb 3, 2016

Outline

Outline

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

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

Outline

Interpreting the Correlation

- If r is close to $+1$ or -1, the relationship is strong.
- If r is close to 0, the relationship is weak.

Interpreting the 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.

Interpreting the 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.
- If $r > 0$, the relationship is positive.
- If $r < 0$, the relationship is negative.

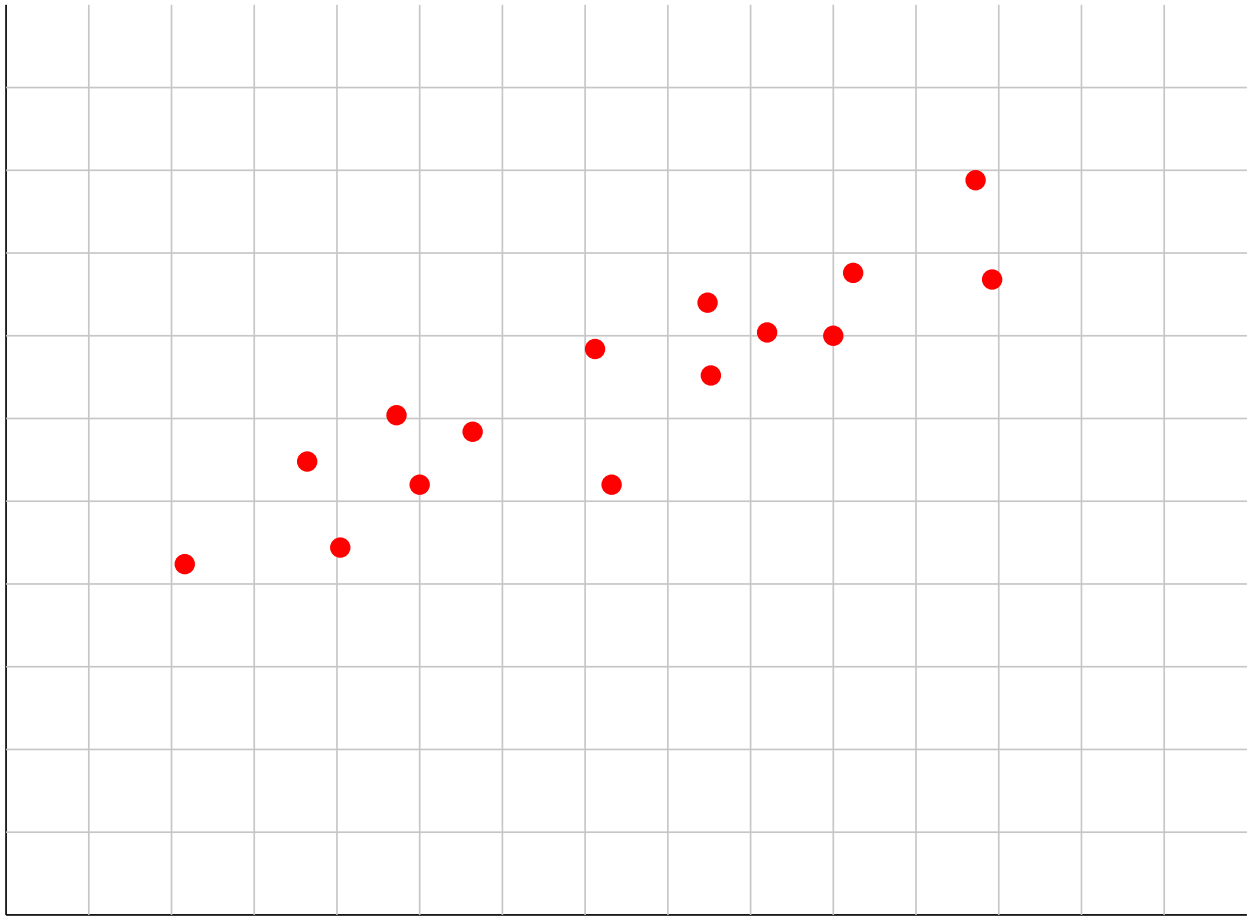
Interpreting the Correlation

- 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.
- If $r > 0$, the relationship is positive.
- If $r < 0$, the relationship is negative.
- In other words, the sign of r indicates the direction of the relationship.

Interpreting the Correlation

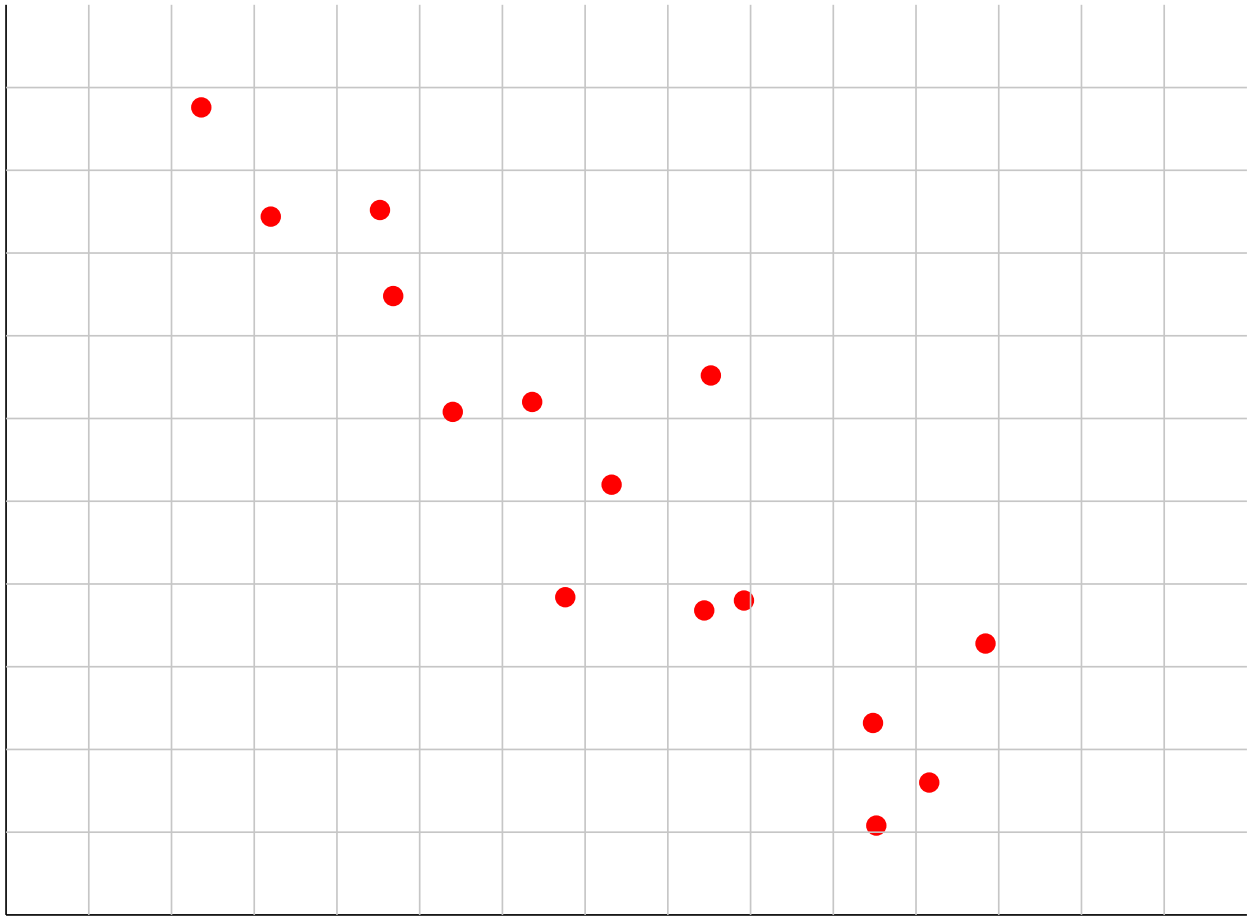
- If r is close to $+1$ or -1, the relationship is strong.
- If r is close to 0, the relationship is weak.
- In other words, $|r|$ measures the strength of the relationship.
- If $r > 0$, the relationship is positive.
- If $r < 0$, the relationship is negative.
- In other words, the sign of r indicates the direction of the relationship.
- The correlation makes no distinction between explanatory and response variables.

Strong Positive Linear Association



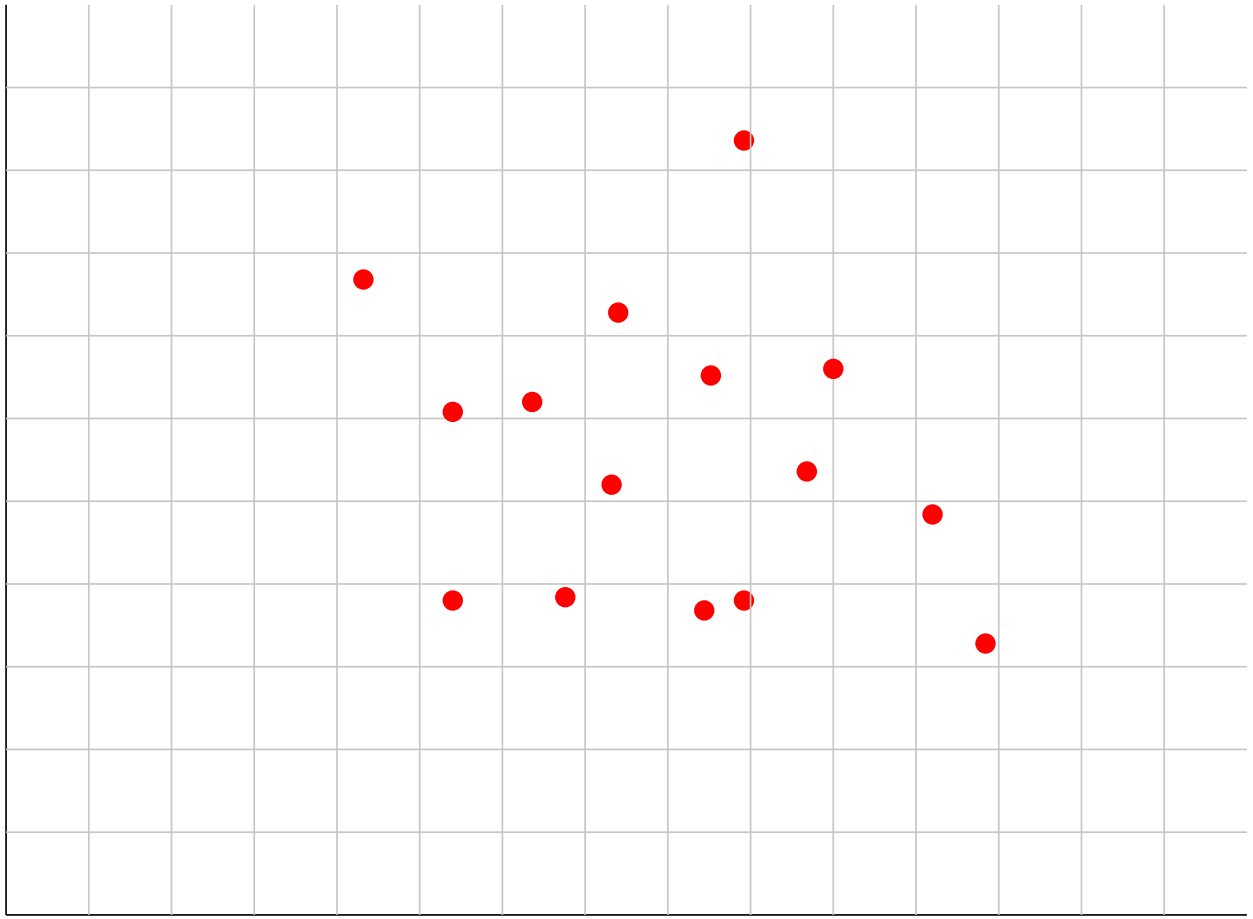
r is close to $+1$

Strong Positive Linear Association



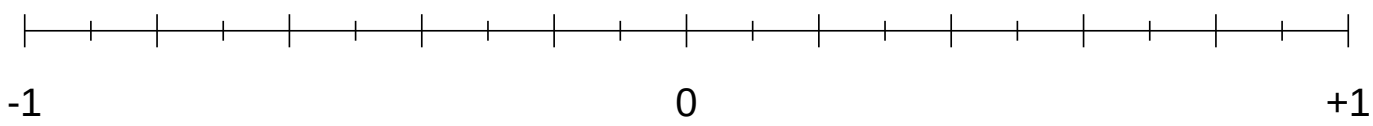
r is close to -1

Strong Positive Linear Association

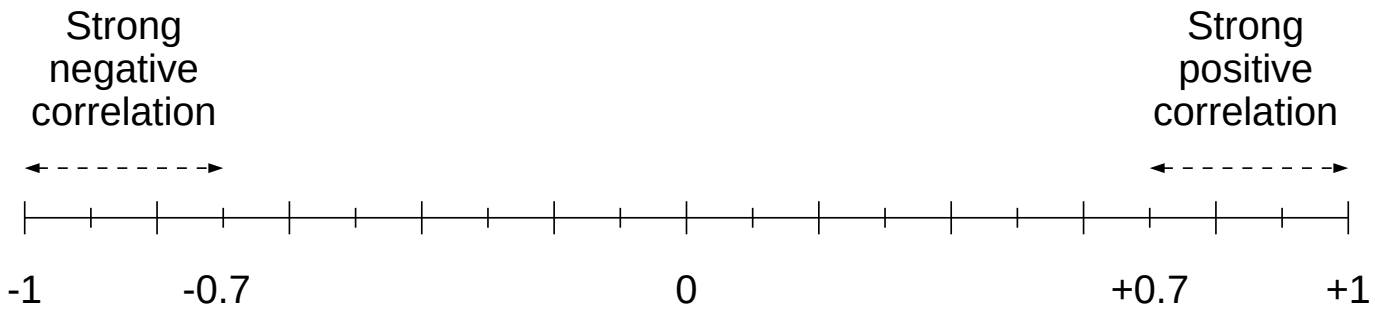


r is close to 0

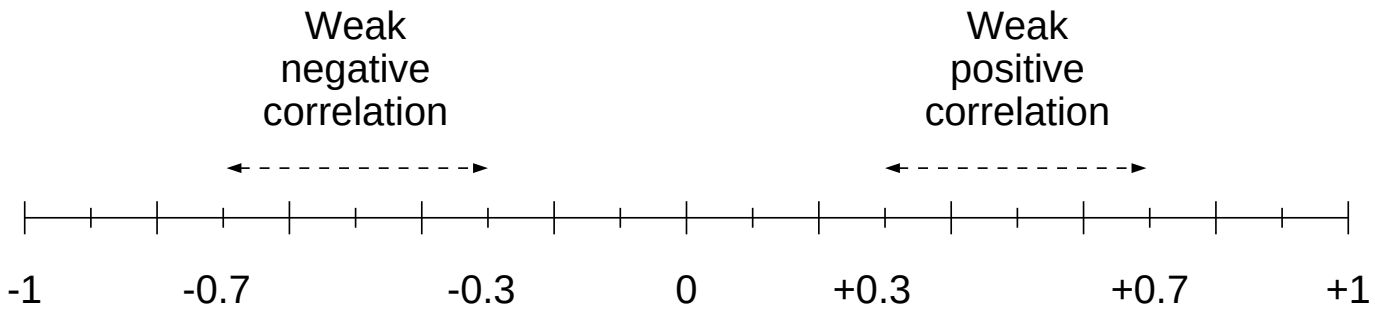
Strong Positive Linear Association



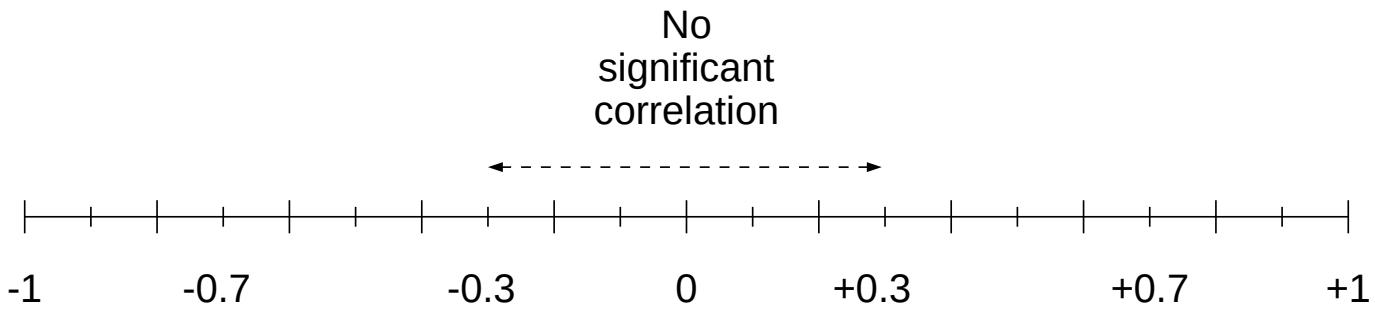
Strong Positive Linear Association



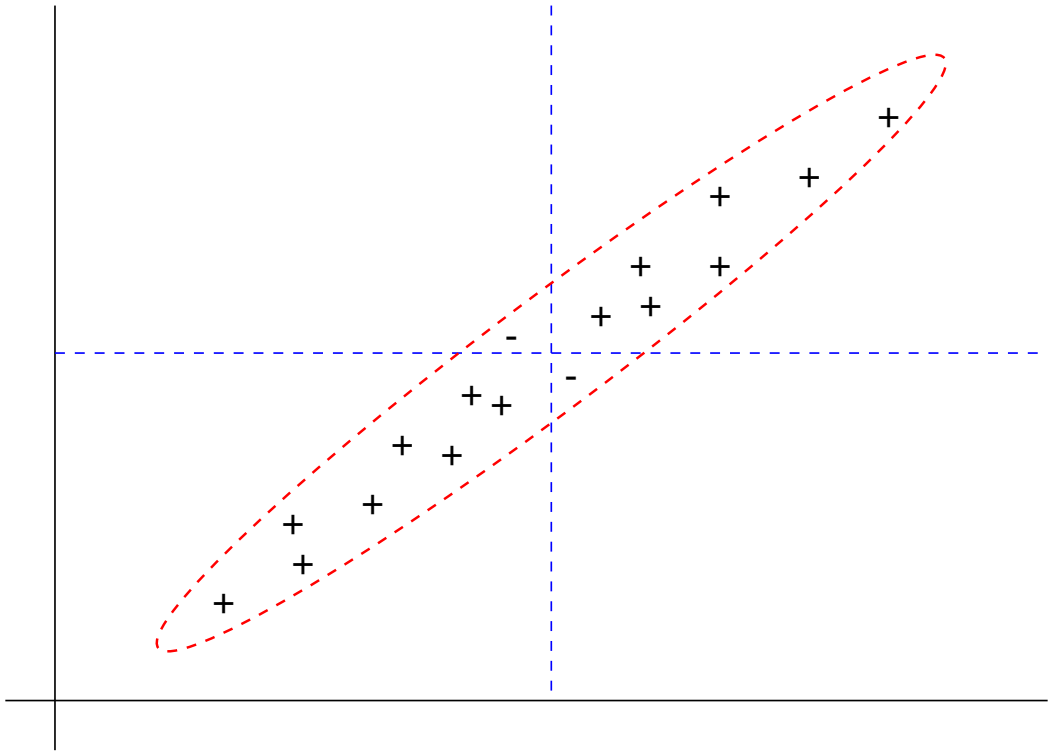
Strong Positive Linear Association



Strong Positive Linear Association

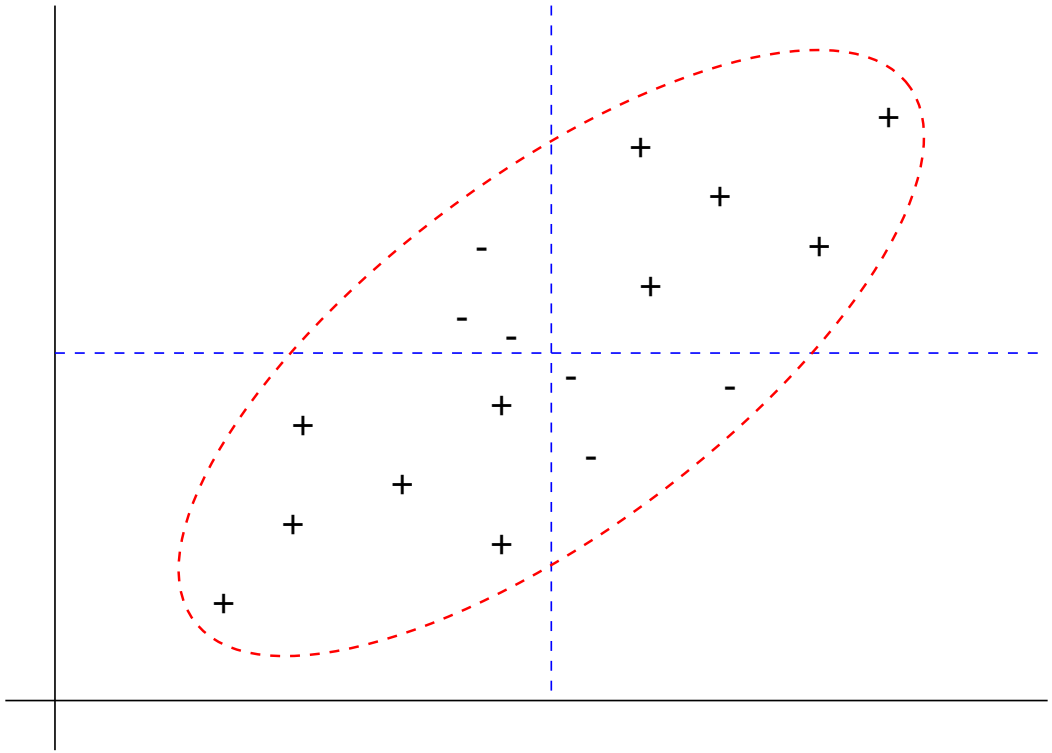


Why the Formula Works



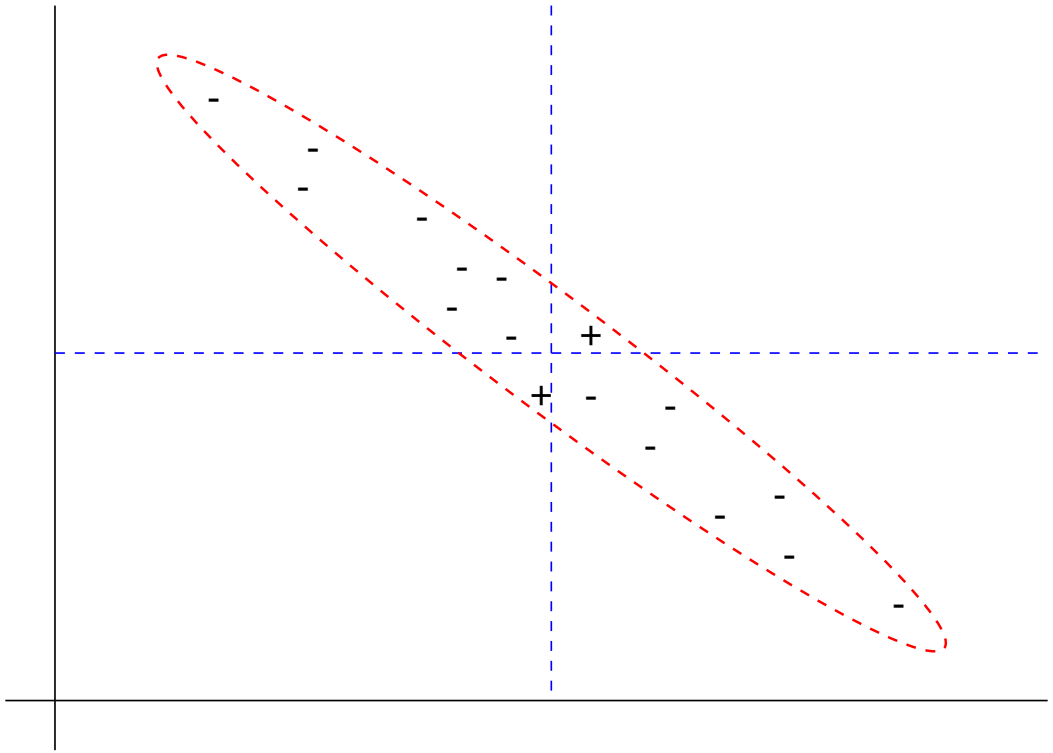
Almost all positive products

Why the Formula Works



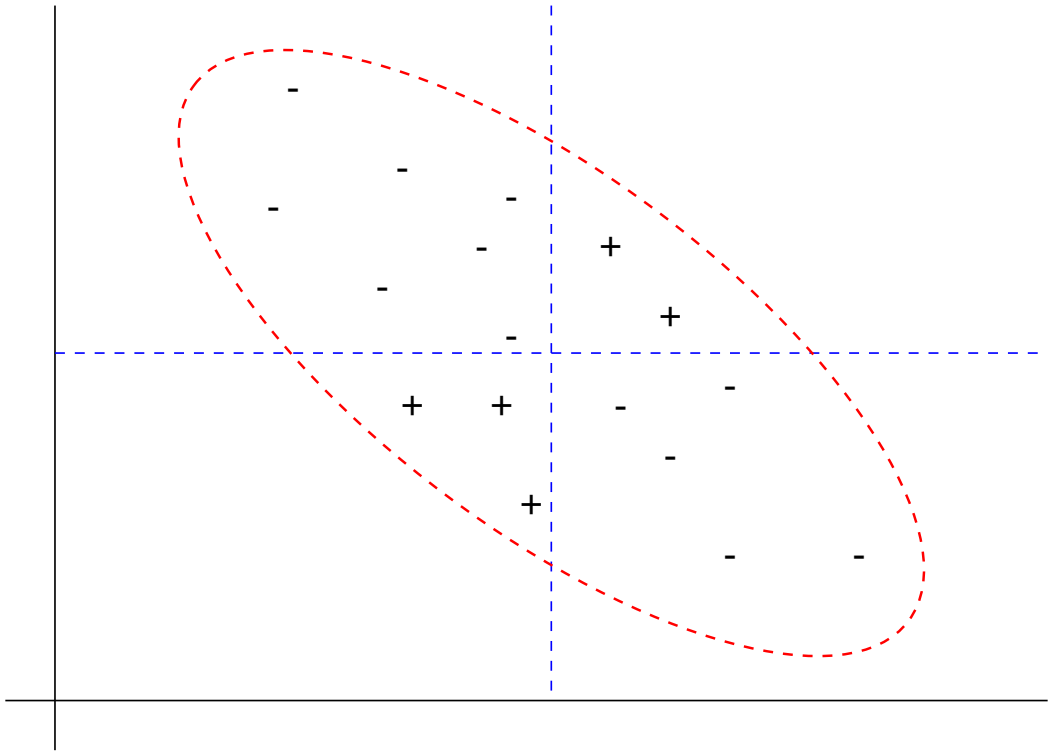
Mostly positive products

Why the Formula Works



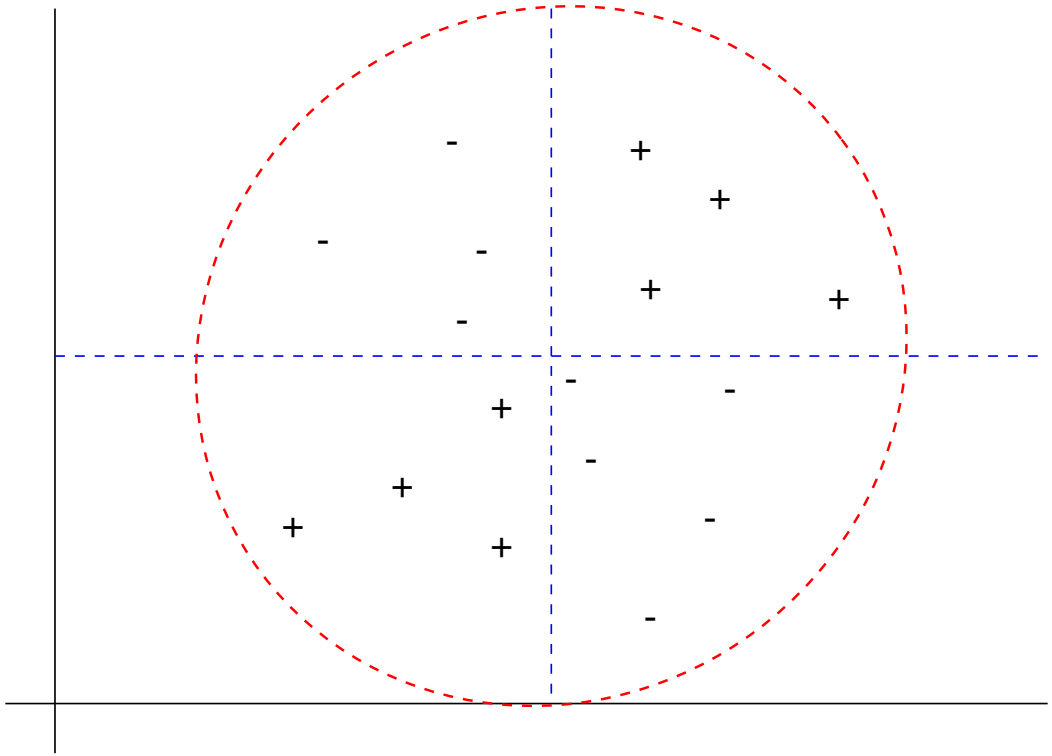
Almost all negative products

Why the Formula Works



Mostly negative products

Why the Formula Works



Even mix of positive and negative products

Outline

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.

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:

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

- We may simplify the formula:

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

- We may simplify the formula:

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

Outline

Practice

Practice

- Find the correlation of the data

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

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40					
5	80					
7	160					
9	180					
16	240					
40	700					

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40	-5				
5	80	-3				
7	160	-1				
9	180	1				
16	240	8				
40	700	0				

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40	-5	-100			
5	80	-3	-60			
7	160	-1	20			
9	180	1	40			
16	240	8	100			
40	700	0	0			

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40	-5	-100	25		
5	80	-3	-60	9		
7	160	-1	20	1		
9	180	1	40	1		
16	240	8	100	64		
40	700	0	0	100		

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40	-5	-100	25	10000	
5	80	-3	-60	9	3600	
7	160	-1	20	1	400	
9	180	1	40	1	1600	
16	240	8	100	64	10000	
40	700	0	0	100	25600	

Practice

Practice

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
3	40	-5	-100	25	10000	500
5	80	-3	-60	9	3600	180
7	160	-1	20	1	400	-20
9	180	1	40	1	1600	40
16	240	8	100	64	10000	800
40	700	0	0	100	25600	1500

Outline

Using the TI-83

- The TI-83 will calculate r if we put the x values and the y values into lists.

Correlation on the TI-83

Correlation on the TI-83

- Put the x values in $\texttt{L}_1$.
- Put the y values in $\texttt{L}_2$.
- Select `STAT > CALC > LinReg(a+bx)` (item #8).
- Press `Enter`. `LinReg` appears in the display, followed by several numbers.
- The last number `r` is the correlation.

Correlation on the TI-83

Correlation on the TI-83

- If you did not see `r`, then you must turn `Diagnostic` on (once and for all).
- To do so, press `CATALOG` and select `DiagnosticOn`, and press `ENTER` twice.
- Now the value of `r` should appear when you use `LinReg`.

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			

Find the correlation.

Outline

Assignment

Assignment

- Read Sections 4.5, 4.6.
- Apply Your Knowledge: 1, 3, 4, 6.
- Check Your Skills: 14, 15, 19, 20.
- Exercises: 24, 26, 27, 28, 29.