

Lecture 4: Correlation

Dr. Julie VanDusky-Allen¹

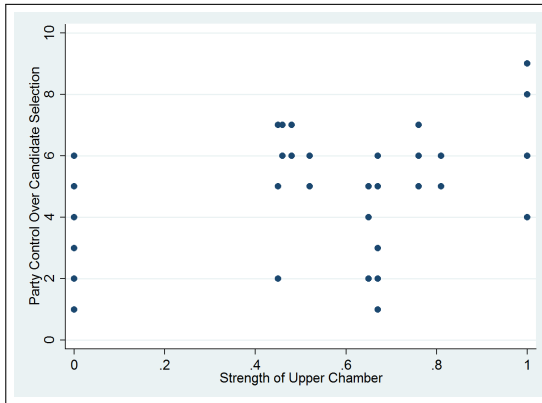
¹Boise State University

Slides prepared for the Political Science MA Program

Objectives

- Understand the intuition behind a Pearson's correlation coefficient
- Calculate a Pearson's correlation coefficient
- Interpret a Pearson's correlation coefficient
- Know how to run a Pearson's correlation analysis in STATA
- Understand the limitations of using a Pearson's correlation coefficient

Scatterplot



What It Tells Us

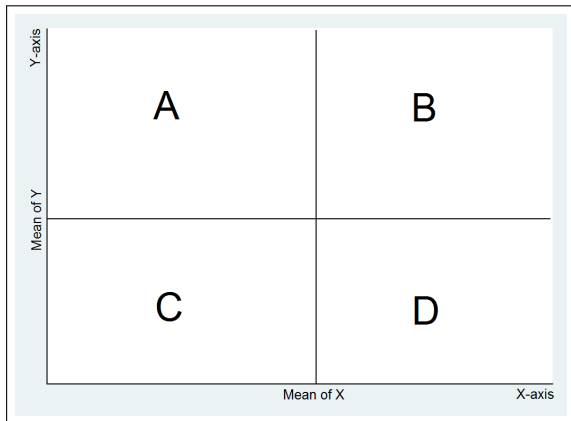
- Direction
- Magnitude
- Values
 - Ranges from -1 to 1
 - -1 = Strong Negative
 - 1 = Strong Positive
 - 0 = Neutral

Equation

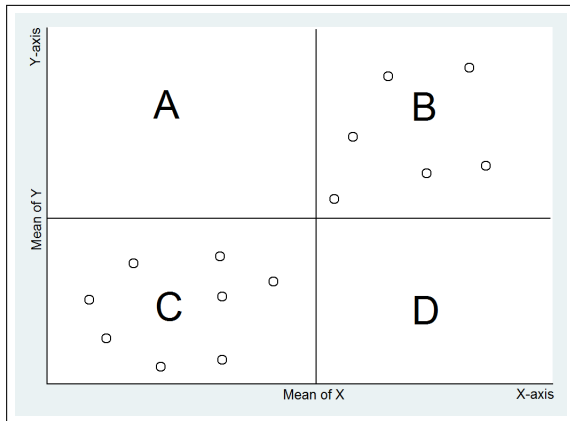
$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n-1}$$

What the Math is Doing

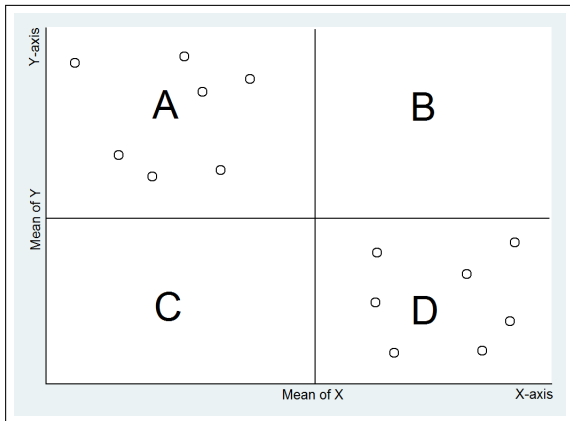
- Each X with its corresponding Y
- Deviation from the mean
- On average, does each X deviate from its mean in a similar pattern to its corresponding Y deviating from its mean?



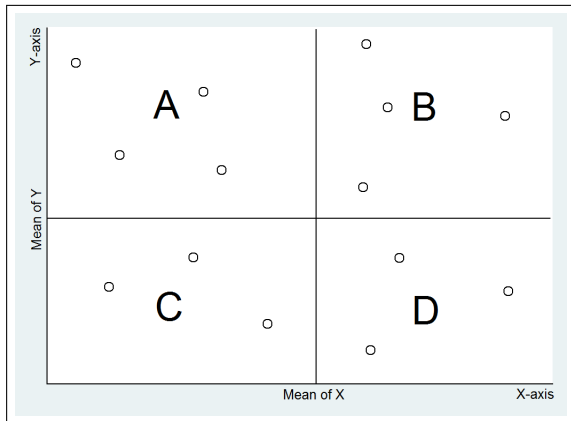
Positive Relationship



Negative Relationship



No Relationship



Hypothetical Example

- Variable X Mean = 50
- Variable Y Mean = 10
- Observation 1: $X = 70$, $Y = 20$
- Observation 2: $X = 30$, $Y = 5$
- Positive relationship

Hypothetical Example

- Variable X Mean = 50
- Variable Y Mean = 10
- Observation 1: $X = 70$, $Y = 5$
- Observation 2: $X = 30$, $Y = 20$
- Negative relationship

Hypothetical Example

- Variable X Mean = 50
- Variable Y Mean = 10
- Observation 1: $X = 70, Y = 20$
- Observation 2: $X = 30, Y = 20$
- No relationship

Standardization

- Denominator
- Individual deviation versus average deviation

Step 1: Calculate the Means

Table :

Obs	Candidate Selection	Strength
1	6	0.67
2	2	0.65
3	4	1
4	6	0.76
\bar{X}	0.5019757	
\bar{Y}	4.213779	

Step 2: Calculate the Standard Deviation

Table :

Obs	Candidate Selection	Strength
1	6	0.67
2	2	0.65
3	4	1
4	6	0.76
\bar{X}	0.5019757	
\bar{Y}	4.213779	
s_X	0.3448112	
s_Y	1.854738	

Step 3: Subtract the Mean of X from each X

Table :

Obs	Candidate Selection	Strength	$x_i - \bar{X}$
1	6	0.67	$0.67 - 0.5019757 = 0.1680243$
2	2	0.65	$0.65 - 0.5019757 = 0.1480243$
3	4	1	$1 - 0.5019757 = 0.4980243$
4	6	0.76	$0.76 - 0.5019757 = 0.2580243$
\bar{X}	0.5019757		
\bar{Y}	4.213779		
s_X	0.3448112		
s_Y	1.854738		

Step 4: Divide the values from Step 3 by s_X

Table :

Obs	Candidate Selection	Strength	$x_j - \bar{X}$	Divide by s_X
1	6	0.67	0.1680243	$0.1680243/0.3448112 = 0.4872937$
2	2	0.65	0.1480243	$0.1480243/0.3448112 = 0.4292908$
3	4	1	0.4980243	$0.4980243/0.3448112 = 1.444339$
4	6	0.76	0.2580243	$0.2580243/0.3448112 = 0.748306$
\bar{X}	0.5019757			
\bar{Y}	4.213779			
s_X	0.3448112			
s_Y	1.854738			

Step 5: Subtract the Mean of Y from each Y

Table :

Obs	Candidate Selection	Strength	$x_i - \bar{X}$	Divide by s_X	$y_i - \bar{Y}$
1	6	0.67	0.1680243	0.4872937	6-4.213779 = 1.786221
2	2	0.65	0.1480243	0.4292908	2-4.213779 = -2.21378
3	4	1	0.4980243	1.444339	4-4.213779 = -0.21378
4	6	0.76	0.2580243	0.748306	6-4.213779 = 1.786221
\bar{X}	0.5019757				
\bar{Y}	4.213779				
s_X	0.3448112				
s_Y	1.854738				

Step 6: Divide the values from Step 5 by s_Y

Table :

Obs	Candidate Selection	Strength	$x_i - \bar{X}$	Divide by s_X	$y_i - \bar{Y}$	Divide by s_Y
1	6	0.67	0.1680243	0.4872937	1.786221	$1.786221/1.854738 = 0.963058$
2	2	0.65	0.1480243	0.4292908	-2.213779	$-2.213779/1.854738 = -1.193580$
3	4	1	0.4980243	1.444339	-0.213779	$-0.213779/1.854738 = -0.115261$
4	6	0.76	0.2580243	0.748306	1.786221	$1.786221/1.854738 = 0.963058$
\bar{X}	0.5019757					
\bar{Y}	4.213779					
s_X	0.3448112					
s_Y	1.854738					

Step 7: Multiply the Corresponding Values from Steps 4 and 6

Table :

Obs	Candidate Selection	Strength	$x_i - \bar{X}$	Divide by s_x	$y_i - \bar{Y}$	Divide by s_y	4*6
1	6	0.67	0.1680243	0.4872937	1.786221	0.963058	$0.4872937 * 0.963058 = 0.469292$
2	2	0.65	0.1480243	0.4292908	-2.213779	-1.193580	$0.4292908 * -1.193580 = -0.51239$
3	4	1	0.4980243	1.444339	-0.213779	-0.115261	$1.4443339 * -0.115261 = -0.16648$
4	6	0.76	0.2580243	0.748306	1.786221	0.963058	$0.748306 * 0.963058 = 0.720662$
\bar{X}	0.5019757						
\bar{Y}	4.213779						
s_x	0.3448112						
s_y	1.854738						

Step 8: Sum the Values from Step 7

Table :

Obs	Candidate Selection	Strength	$x_i - \bar{X}$	Divide by s_X	$y_i - \bar{Y}$	Divide by s_Y	4*6
1	6	0.67	0.1680243	0.4872937	1.786221	0.9630583	0.469292
2	2	0.65	0.1480243	0.4292908	-2.213779	-1.193580	-0.51239
3	4	1	0.4980243	1.444339	-0.213779	-0.115261	-0.16648
4	6	0.76	0.2580243	0.748306	1.786221	0.963058	0.720662
\bar{X}	0.5019757						
\bar{Y}	4.213779						
s_X	0.3448112						
s_Y	1.854738						
Sum of Step 7 (note) =	377.248071						

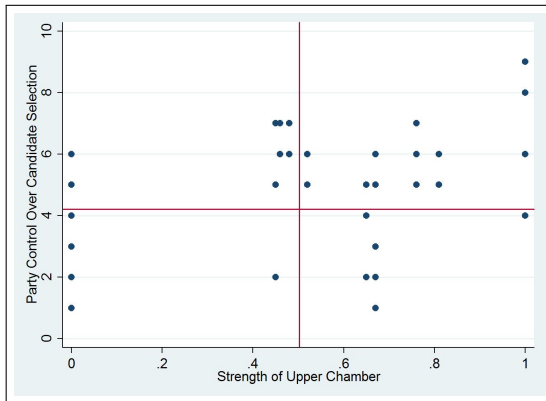
Step 9: Divide the Value from Step 8 by N-1

- $N = 987$
- $N-1 = 986$
- $\frac{377.248071}{986} = 0.383$
- $r = 0.383$

What It Tells Us

- Direction and magnitude
- Values
 - Ranges from -1 to 1
 - -1 = Strong Negative
 - 1 = Strong Positive
 - 0 = Neutral
- $r = 0.383$

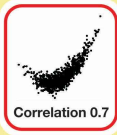
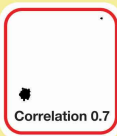
Scatterplot



STATA

- Type “corr varname1 varname2” into the command box and press enter
- The output appears in the output window
- The cell that is at the intersection of both variables is the correlation coefficient for those variables

*Pick a Box!
Are they all the same?*



SECTION ON
STATISTICAL
GRAPHICS



AMERICAN
STATISTICAL
ASSOCIATION

<http://www.amstat-online.org/sections/graphics/>

Substantive Effects

- Specific impact
 - As X changes by 1, how much does Y change?

Relevant XKCD

