

Basics of Correlation

The correlation coefficient can range in value from -1 to +1, and tells you two things about the linear relationships between two variables:

1. **Strength** - The larger the absolute value of the coefficient, the stronger the linear relationship between the variables.
 - An absolute value of one (1.0) indicates a perfect linear relationship, and a value of zero (0) indicates the absence of a linear relationship.
 - Whether a correlation coefficient is interpreted as a weak, moderate, or strong correlation depends on your objectives and requirements.
 - We tend to pay attention if the value is above .40 or so.

Basics of Correlation

2. Direction - The **sign of the coefficient** indicates the direction of the relationship.

- If both variables tend to increase or decrease together, the coefficient is positive.
- If one variable tends to increase as the other decreases, the coefficient is negative.

We can compute the **line of best fit** – the line that minimizes the errors in prediction

Correlation

- Two types:
 - Pearsons r -- Called the correlation coefficient or more fully, the product moment correlation coefficient. This is used to measure the relationship between two sets of interval or ratio data
 - Spearmans rho – interpreted the same as Pearson's r, but is used to measure **ranked data**.

- Pearson's r, Where

N = number of observations

X = X variable, Y = Y variable

Thus,

$$r = \frac{\sum_{XY} - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N} \right)} \sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{N} \right)}}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N} \right)} \sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{N} \right)}}$$

Pearsons r

Subject	Linguistic aptitude Final Score			Values Needed to Calculate r			
	(X)	(Y)	XY	X ²	Y ²	N	Value
1	5	5	25	25	25		12
2	10	20	200	100	400	Sum of X	74
3	6	4	24	36	16	Sum of X squared	5476
4	8	15	120	64	225	Sum of Y	134
5	4	11	44	16	121	Sum of Y squared	17956
6	4	9	36	16	81	Sum of XY	942
7	3	12	36	9	144	Sum of X ²	536
8	10	18	180	100	324	Sum of Y ²	1874
9	2	7	14	4	49		
10	6	2	12	36	4		
11	7	14	98	49	196		
12	9	17	153	81	289		
Sums ->	74	134	942	536	1874	r =	0.666831

There is a Correlation function in Excel

The screenshot shows a Microsoft Excel spreadsheet titled "CorrelationExercises [Compatibility Mode] - Microsoft Excel". The Data tab is selected in the ribbon. A Data Analysis dialog box is open, showing various statistical tools. The "Correlation" option is highlighted.

Data Analysis Dialog Box:

- Analysis Tools:
 - Anova: Single Factor
 - Anova: Two-Factor With Replication
 - Anova: Two-Factor Without Replication
 - Correlation** (highlighted)
 - Covariance
 - Descriptive Statistics
 - Exponential Smoothing
 - F-Test Two-Sample for Variances
 - Fourier Analysis
 - Histogram
- OK button
- Cancel button
- Help button

Excel Worksheet Data:

	A	B	C	D	E	F	G	H	I	J	K
1	Pearson's r		Linguistic aptitude (X)	Final Score (Y)	XY	X ²	Y ²	Values Needed to Calculate r			
2	Subject		(X)	(Y)	25	25	25		Value		
3	1		5	5	200	100	400	N	12	Linguistic aptitude (X)	
4	2		10	20				Sum of X	74	Final Score (Y)	
5	3		6	4							
6	4		8	15							
7	5		4	11							
8	6		4	9							
9	7		3	12							
10	8		10	18							
11	9		2	7							
12	10		6	2							
13	11		7	14							
14	12		9	17							
15	Sums ->		74	134	942	536	1874				
16											
17											
18	Spearman's rho										
19	Subject	Linguistic aptitude	(Y) ranked	d = (X-Y)	d ²			Calculate rho	Value		
20	1		8	10	-2	4		N	12	Regression	

Bottom Taskbar:

- Sheet1 / Sheet2 / CorrelationExercises
- Ready
- Start button
- Icons for Internet Explorer, Outlook, and File Explorer
- Taskbar buttons for CHI Square, CorrelationE..., Finishing Lectu..., AnalyticBasic..., Spearman's ra..., and a clock showing 12:01 PM

There is a Correlation function in Excel

Linguistic aptitude (X)	Final Score (Y)
5	5
10	20
6	4
8	15
4	11
4	9
3	12
10	18
2	7
6	2
7	14
9	17
74	134

Linguistic aptitude (X)	Final Score (Y)
Linguistic aptitude (X)	1
Final Score (Y)	0.666830675

Spearman's rho

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Spearman's rho

Linguistic aptitude Final Score					Values Needed to Calculate rho	
Subject	(X) ranked	(Y) ranked	d = (X-Y)	d ²	N	Value
1	8	10	-2	4		12
2	1	1	0	0		
3	6	11	-5	25		
4	4	4	0	0		
5	9	7	2	4		
6	9	8	1	1		
7	11	6	5	25		
8	1	2	-1	1		
9	12	9	3	9		
10	6	12	-6	36		
11	5	5	0	0		
12	3	3	0	0		
Sums ->				105		

rho= 0.63287