

Opinkirjolle koontu Visajaani Salonen (2018).

Pearsonin korrelaatiokerroin r lasketaan seuraavasti:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$$

missä

n = lukuparien x_i, y_i lukumäärä,

σ_x, σ_y ovat keskihajonnat muuttujille x ja y ,

\bar{x} ja \bar{y} ovat keskiarvot muuttujille x ja y .

Vaikka korrelaatiokertoimen laskemiseen kaava vaikuttaa monimutkaiselta, usein korrelaatiokerroin lasketaan valmiiden funktioiden avulla. Microsoft Excel-ohjelmiston funktion CORREL (suomenkielinen versio KORRELAATIO) avulla ei tarvitse rakentaa erikseen kaavoja korrelaatiokertoimen laskemiseksi.

A	B	C	D	E
	ääni ja lippu	ääni		
1				
2	0,36	0,621		
3	0,36	0,223		
4	0,63	0,296		
5	0,43	0,567		
6	0,32	0,271		
7	0,37	0,466		
8	0,15	0,271		
9	0,29	0,299		
10	0,35	0,284		
11	0,38	0,464		
12	0,10	0,338		
13	0,13	0,705		
14	0,67	0,137		
15	0,06	0,387		
16	0,66	0,697		
17	0,42	0,71		
18	0,31	0,47		
19	0,73	0,204		
20	0,44	0,18		
21	0,34	0,786		
22	0,16	0,603		
23	0,15	0,446		
24	0,14	0,233		
25	0,71	0,269		
26	0,36	0,51		
27	0,08	0,571		
28	0,77	0,09		
29	0,06	0,446		
30	0,08	0,574		
31	0,09	0,28		
32	0,45	0,765		
33	0,41	0,262		
34	0,53	0,77		
35	0,59	0,798		
36	0,11	0,583		
37	0,79	0,1		
38	0,55	0,147		
39	0,66	0,28		
40	0,24	0,423		
41	0,37	0,42		
42	Keskiarvo	0,37	0,42	
43	Varianssi	0,047292594	0,041902886	
44	T-testi	0,320207551		
45				
46				
47	=CORREL(B2:B41;C2:C41)			
48	CORREL(array1; array2)			

Korrelaation merkitsevyys voidaan testata Microsoft Excel-ohjelmiston T.DIST funktion avulla, kunhan ensin lasketaan t-arvo korrelaatiolle kaavalla:

$$t = \frac{r * \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

Kaavassa r on korrelaatio ja n on havaintojen lukumäärä. Microsoft Excel-ohjelmistoon kaava voidaan toteuttaa seuraavasti:

Opinkirjolle koonnut Visajaani Salonen (2018).

CONCATENATE...						G49			
A	B	C	D	E	F	A	B	C	D
1	ääni ja lippu	ääni				1	ääni ja lippu	ääni	
2	0,36	0,621				2	0,36	0,621	
3	0,36	0,223				3	0,36	0,223	
4	0,63	0,296				4	0,63	0,296	
5	0,43	0,567				5	0,43	0,567	
6	0,32	0,271				6	0,32	0,271	
7	0,37	0,466				7	0,37	0,466	
8	0,15	0,271				8	0,15	0,271	
9	0,29	0,299				9	0,29	0,299	
10	0,35	0,284				10	0,35	0,284	
11	0,38	0,464				11	0,38	0,464	
12	0,10	0,338				12	0,10	0,338	
13	0,13	0,705				13	0,13	0,705	
14	0,67	0,137				14	0,67	0,137	
15	0,06	0,387				15	0,06	0,387	
16	0,66	0,697				16	0,66	0,697	
17	0,42	0,71				17	0,42	0,71	
18	0,31	0,47				18	0,31	0,47	
19	0,73	0,204				19	0,73	0,204	
20	0,44	0,18				20	0,44	0,18	
21	0,34	0,786				21	0,34	0,786	
22	0,16	0,603				22	0,16	0,603	
23	0,15	0,446				23	0,15	0,446	
24	0,14	0,233				24	0,14	0,233	
25	0,71	0,269				25	0,71	0,269	
26	0,36	0,51				26	0,36	0,51	
27	0,08	0,571				27	0,08	0,571	
28	0,77	0,09				28	0,77	0,09	
29	0,06	0,446				29	0,06	0,446	
30	0,08	0,574				30	0,08	0,574	
31	0,09	0,28				31	0,09	0,28	
32	0,45	0,765				32	0,45	0,765	
33	0,41	0,262				33	0,41	0,262	
34	0,53	0,77				34	0,53	0,77	
35	0,59	0,798				35	0,59	0,798	
36	0,11	0,583				36	0,11	0,583	
37	0,79	0,1				37	0,79	0,1	
38	0,55	0,147				38	0,55	0,147	
39	0,66	0,28				39	0,66	0,28	
40	0,24	0,423				40	0,24	0,423	
41	0,37	0,42				41	0,37	0,42	
42	Keskiarvo	0,37	0,42			42	Keskiarvo	0,37	0,42
43	Varianssi	0,047292594	0,041902886			43	Varianssi	0,047292594	0,041902886
44	T-testi	0,320207551				44	T-testi	0,320207551	
45	Otos N	40				45	Otos N	40	
46						46			
47	Korrelaati	-0,254562568				47	Korrelaati	-0,254562568	
48	=(B47*SQRT(B45-2))/(SQRT(1-B47^2))					48	t-arvo	-1,622686339	
49						49			
50	SQRT(number)					50			
51						51			

Tämän jälkeen T.DIST(t-arvo;havainnot-2;suuntaisuus)-funktiolla laskemme p-arvon korrelaatiolle. Ensimmäisenä syötämme lasketun t-arvon. Toiseen kohtaan sijoitamme havainnot, joista vähennämme 2. Viimeiseen määrittelykohtaan laitamme suuntaisuudeksi 2(kaksisuuntainen), kun vastaavasti yksisuuntaiselle testille olisi määrite 1.

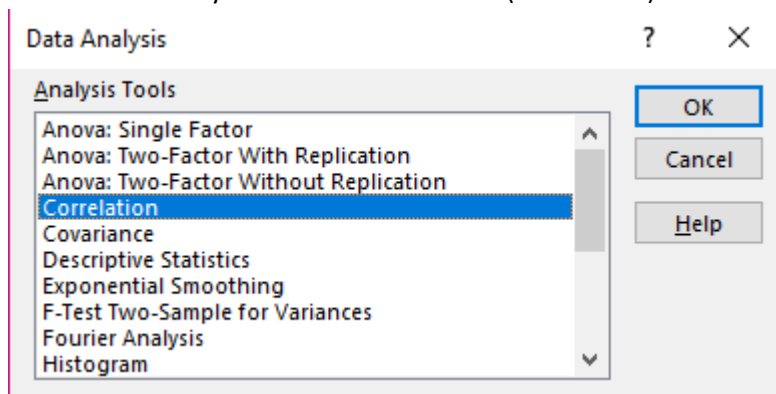
Opinkirjolle koonnut Visajaani Salonen (2018).

Nyt voimme raportoida korrelaation taulukkona:

	<i>ääni ja lippu</i>	<i>ääni</i>
<i>ääni ja lippu</i>	1	
<i>ääni</i>	-0,25456	1

Taulukon saa helposti muodostettua suoraan Analysis ToolPak-lisäosan avulla.

1. Valitse Data Analysis valikosta Correlation(Korrelaatio)



2. Valitse Input Range() kohtaan taulukon ruudut, jotka sisältävät alkuperäiset arvot sekä otsikot. Klikkaa valituksi Labels in First Row, jotta saat taulukkoon nimet automaattisesti mukaan.

Opinkirjolle koonnut Visajaani Salonen (2018).

The screenshot shows the 'Correlation' dialog box in Microsoft Excel. The 'Input' section has 'Input Range' set to '\$B\$1:\$C\$41', 'Grouped By' set to 'Columns', and 'Labels in First Row' checked. The 'Output options' section has 'New Worksheet Ply' selected. Below the dialog box, a data table is visible with columns B and C highlighted. The table contains 41 rows of data, with the first row containing the labels 'ääni ja lippu' and 'ääni'. The last row (row 42) shows the average values for columns B and C as 0.37 and 0.42, respectively.

	B	C
1	ääni ja lippu	ääni
2	0,36	0,621
3	0,36	0,223
4	0,63	0,296
5	0,43	0,567
6	0,32	0,271
7	0,37	0,466
8	0,15	0,271
9	0,29	0,299
10	0,35	0,284
11	0,38	0,464
12	0,10	0,338
13	0,13	0,705
14	0,67	0,137
15	0,06	0,387
16	0,66	0,697
17	0,42	0,71
18	0,31	0,47
19	0,73	0,204
20	0,44	0,18
21	0,34	0,786
22	0,16	0,603
23	0,15	0,446
24	0,14	0,233
25	0,71	0,269
26	0,36	0,51
27	0,08	0,571
28	0,77	0,09
29	0,06	0,446
30	0,08	0,574
31	0,09	0,28
32	0,45	0,765
33	0,41	0,262
34	0,53	0,77
35	0,59	0,798
36	0,11	0,583
37	0,79	0,1
38	0,55	0,147
39	0,66	0,28
40	0,24	0,423
41	0,37	0,42
42 Keskiarvo	0,37	0,42

3. Paina lopuksi ok ja tulokset avautuvat uuteen Microsoft Excel-välilehteen

The screenshot shows the result of the correlation analysis in a new Microsoft Excel worksheet. The data table is displayed with the following values:

	A	B	C
1		ääni ja lippu	ääni
2	ääni ja lip	1	
3	ääni	-0,25456	1
4			
5			
6			

Koska tässä tapauksessa ei ollut merkitsevää korrelaatiota, ei kuvioon tarvitse merkitä korrelaation merkitsevyyksiä näkyviin. Yleensä raportoidaan vain merkitsevät korrelaatiot taulukosta merkitsemällä

Opinkirjolle koonnut Visajaani Salonen (2018).

merkitsevyys asteen mukaan tähdillä taulukkoon luvun perään (* on $p < 0,05$, ** on $p < 0,01$ ja *** on $p < 0,001$). Tässä tapauksessa kirjallisesti voitaisiin raportoida seuraavasti:

”Starttipistoolin sekä lipun kanssa lähetetyn juoksukilpailun ja pelkästään starttipistoolin avulla lähetetyn juoksukilpailun välillä oli negatiivinen korrelaatio, joka ei kuitenkaan ollut tilastollisesti merkitsevä ($r = -0,255$ ja $p = 0,056$)”