

Biomeetria praks 5

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excelis* oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks5' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks5' ülemisse vasakusse nurka.

Ülesanne 1.

Kas tudengite pikkus ja jalanumber on seotud? Uurige seost *MS Exceli* funktsioonide abil.

- Leidke tunnuste 'PIKKUS' ja 'JALANR' vaheline lineaarne korrelatsioonikordaja;
- kirjeldage pikkuse ja jalanumbri vahelist seost leitud kordaja alusel;
- testige seose statistilist olulisust:
 - sõnastage null- ja alternatiivne hüpotees (pange need töölehele kirja),
 - kontrollige nende kehtimist (leidke vaatluspaaride arv n , teststatistik t ja nende alusel olulisuse tõenäosus p),
 - pange kirja lõppjärelendus.

Ülesanne 2.

Illustreerige tunnuste 'PIKKUS' ja 'JALANR' vahelist seost hajuvus- e punktdiagrammiga (inglise keeles *scatter plot*).

Ülesanne 3.

- Kasutades statistikaprotseduuri *Correlation (Data-sakk -> Data analysis...)* leidke korruga kõigi andmestikku kuuluvate pidevate arvtunnuste (pikkus – jalanumber) vahelised lineaarsed korrelatsioonikordajad.
- Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?
- Kirjeldage mõnda seost (pange vastavad laused ka kirja).

Ülesande 1 tööjuhend

1. Et MS Exceli funktsioonide tulemuseks on enamasti vaid üks kommenteerimata väärtus, on enne millegi välja arvutamist soovitatav kirja panna, mis see on, mida arvutatakse.

Näiteks antud juhul on eesmärgiks leida tudengite pikkuse ja jalanumbri vaheline lineaarne korrelatsioonikordaja – MS Exceli töölehele võiks siis trükkida


- 'Pikkuse ja jalanumbri vaheline lineaarne korrelatsioonikordaja'
- või lühemalt 'r(Pikkus;Jalanr)', sest lineaarset korrelatsioonikordajat tähistatakse enamasti tähega 'r'.

Seejärel pange kursor tühja lahtrisse, millesse soovite vastavat korrelatsioonikordajat arvutada.

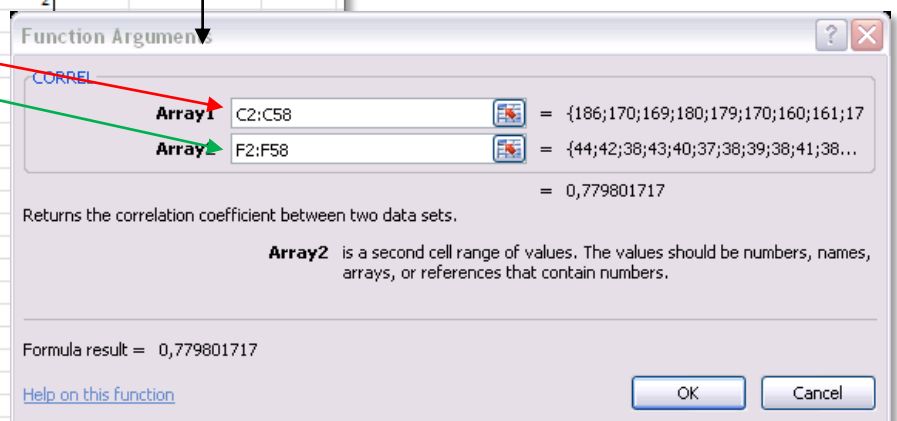
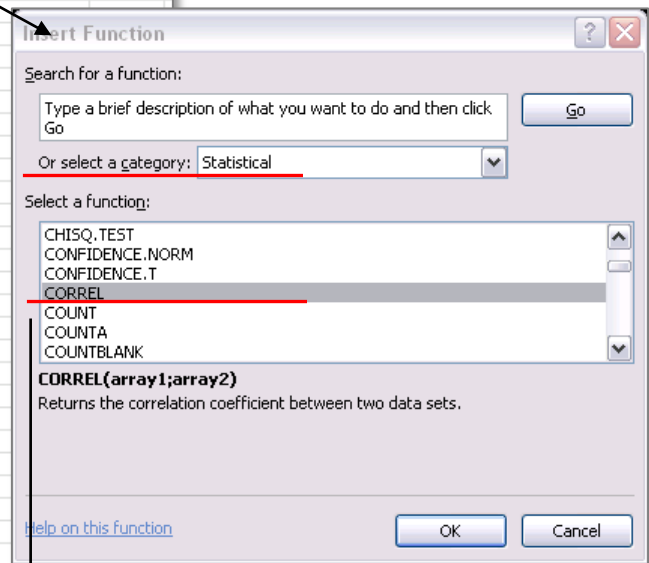
2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

- Kogenumad MS Exceli kasutajad võivad kohe sisestada arvutamiskäsu kujul

=CORREL (C2 : C58 ; F2 : F58)

- Vähem kogenumatel (eelnevast valemist mitte arusaanutel) on soovitatav klikkida nupul  ja jätkata vastavalt joonisele.

	A	B	C	D	E	F	G	T	U	V
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEN		r(Pikkus;Jalanr)	
2	Eesti	M	186	95	59	44	1			
3	Eesti	N	170	85	57	42	6			
4	Eesti	N	169	50	54	38	1			
5	Eesti	M	180	70	56	45	0			
6	Eesti	N	179	72	55	40	1			
7	Eesti	N	170	55	55	37	1			
8	Eesti	N	160	58	55	38	1			
9	Eesti	N	161	57	55	39	1			
10	Eesti	N	171,5	59	57	38	1			
11	Eesti	N	180	63	58	41	2			
12	Eesti	N	168	54	57	38	1			
13	Eesti	N	170	57	52	40	2			
14	Eesti	N	163	61	57,5	39	0			
15	Eesti	M	172	66	54	42	1			
16	Eesti	M	183	73	54,5	44				
17	Eesti	M	185	72	56	44	1			
18	Eesti	M	187	94	59	46	1			
19	Eesti	M	183	83	56	43	3			
20	Eesti	M	190	102	59	46	1			
21	Eesti	M	173	58	55,5	42,5	1			
22	Eesti	N	157	63	55,5	38	1			
23	Eesti	M	180	80	56	43	2			
24	Eesti	M	180	84	60	43	1			
25	Eesti	M	175	87	54	45	3			
26	Eesti	M	181	81	55	43	2			
27	Eesti	M	177	75	54	42	1			
28	Eesti	N	175	60	53	39	1			
29	Eesti	M	185	100	67	45	2			
30	Eesti	N	176	75	56	41	0			
31	Eesti	N	170	100	56	42	1			
32	Eesti	M	179	59	56	44	1			
33	Eesti	M	193	75	55	44	3			
34	Eesti	N	169	60	56	38	2			
35	Eesti	M	185	80	60	43	3			
36	Eesti	N	163	64	57	38	4			
37	Eesti	N	181	74	56	41	2			
38	Eesti	M	191	70	59	46				
39	Eesti	M	160	65	55	42				
40	Eesti	M	173	67	55	42				
41	Eesti	N	172	65	53	37				
42	Eesti	N	173	80	56	40				
43	Eesti	N	167	61	57	38				
44	Eesti	M	185	100	60	46				
45	Eesti	M	182	100		47				
46	Eesti	N	171	64	54,5	39				
47	Eesti	N	173	55	57	39				
48	Eesti	N	174	75	57	40				
49	Eesti	N	162	55	57	39				
50	Eesti	N	176	70	46	39				
51	Eesti	N	168	58	55	38				
52	Eesti	N	171	58	55	39				
53	Eesti	N	172	55	53	38				
54	Eesti	M	176	66	55	43				
55	Eesti	M	188	95	59	44				
56	Eesti	M	188	80	48	48				
57	Eesti	M	180	74	56	42				
58	Eesti	N	173	59	58	40				



3. Kirjeldage, milline on tudengite pikkuste ja jalanumbrite vaheline seos

- kui tugev (nõrk / keskmise tugevusega / tugev),
- kas positiivne või negatiivne (mida see positiivne või negatiivne tähendab?).

NB! See järeldus tuleneb üksnes seose positiivsusest/negatiivsusest! Lünka tuleks kirjutada sõna „suurem“ või „väiksem“.

r(Pikkus;Jalanr)	0,779802		
Pikkuse ja jalanumbri vahel on seos.			
St, et mida suurem on pikkus, seda on keskmiselt ka jalanumber.			
Hüpoteeside paar			
H ₀ : (ehk matemaatiliselt r = 0)			
H ₁ : (ehk matemaatiliselt r ≠ 0)			

4. Pange korrelatsioonikordaja kohta kontrollitav hüpoteeside paar kirja ka teksti kujul.

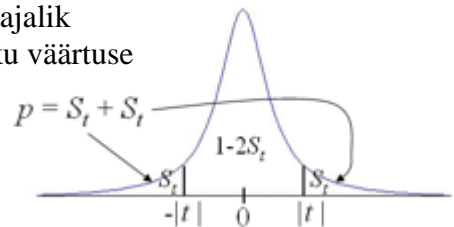
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

Testimaks korrelatsioonikordaja erinevust nullist (st testimaks seose statistilist olulisust) MS Excel'is, tuleb arvutada teststatistik (mis on nullhüpoteesi kehtides t-jaotusega) valemist

$$t = r\sqrt{n-2} / \sqrt{1-r^2} \underset{H_0}{\sim} t_{n-2}.$$

Suurus r selles valemis on arvatud korrelatsioonikordaja väärtus, n aga vaatluspaaride arv (ehk nende tudengite arv, kelle kohta on teada nii pikkuse kui ka jalanumbri väärtus).

Otsuse, kumb hüpoteesidest on õige, vastu võtmiseks vajalik olulisuse tõenäosus p kujutab enesest leitud teststatistiku väärtuse poolt ära lõigatud t-jaotuse sabade osakaalu (joonisel pindalade S_t summa).



Excel 2010-s on p-väärtus leitav funktsiooniga T.DIST.2T (ABS (t) ; n-2).

5. Olulisuse tõenäosuse p leidmiseks vajalikke arvutusi on mugav teostada, kui kõik vajalikud suurused on töölehele kirja pandud.

Näiteks kujul:

- a) Lahtri 'n(Pikkus;Jalanr)' järele tuleks kirjutada nende tudengite arv, kelle andmete alusel on korrelatsioonikordaja arvatud (ehk nende tudengite arv, kelle kohta on teada nii nende pikkus kui ka jalanumber).

r(Pikkus;Jalanr)	0,779802		
Pikkuse ja jalanumbri vahel on tugev positiivne seos.			
St, et mida suurem on pikkus, seda suurem on keskmiselt ka ja			
Hüpoteeside paar			
H ₀ : Pikkus ja jalanumber ei ole seotud (ehk matemaatiliselt r =			
H ₁ : Pikkus ja jalanumber on seotud (ehk matemaatiliselt r ≠ 0)			
n(Pikkus;Jalanr)			
t(Pikkus;Jalanr)			
p(Pikkus;Jalanr)			

- b) Lahtri 't(Pikkus;Jalanr)' järele tuleb sisestada valem teststatistiku absoluutväärtuse arvutamiseks:

	T	U	V	W	X	Y
1		r(Pikkus;Jalanr)	0,779802			
2						
3		Pikkuse ja jalanumbri vahel on tugev positiivne seos.				
4		St, et mida suurem on pikkus, seda suurem on keskmiselt				
5						
6		Hüpoteeside paar				
7		H ₀ : Pikkus ja jalanumber ei ole seotud (ehk matemaatilise				
8		H ₁ : Pikkus ja jalanumber on seotud (ehk matemaatiliselt r				
9						
10						
11		n(Pikkus;Jalanr)	57			
12						
13		t(Pikkus;Jalanr)	=ABS(V1*SQRT(V11-2)/SQRT(1-V1*V1))			
14						
15		p(Pikkus;Jalanr)				

- c) Lahtri 'p(Pikkus;Jalanr)' järele tuleb sisestada funktsioon T.DIST.2T kahe argumentiga:
- teststatistiku absoluutväärtus $|t|$ ja
 - (vaatluspaaride arv) - 2, so vastava t-jaotuse parameeter ($n - 2$).

	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1		r(Pikkus;Jalanr)	0,779802								
2											
3		Pikkuse ja jalanumbri vahel on t									
4		St, et mida suurem on pikkus, se									
5											
6		Hüpoteeside paar									
7		H ₀ : Pikkus ja jalanumber ei ole s									
8		H ₁ : Pikkus ja jalanumber on seo									
9											
10											
11		n(Pikkus;Jalanr)	57								
12											
13		t(Pikkus;Jalanr)	9,237889								
14											
15		p(Pikkus;Jalanr)	=T.DIST.2T(V13;V11-2)								

Function Arguments

T.DIST.2T

X: V13 = 9,237889357

Deg_freedom: V11-2 = 55

Returns the two-tailed Student's t-distribution.

X is the numeric value at which to evaluate the distribution.

Formula result = 8,82306E-13

Help on this function

OK

NB! Vanemates Exceli versioonides puudub funktsioon T.DIST.2T ning kasutada tuleb funktsiooni TDIST. Viimane nõuab kolme argumenti: neist kaks esimest on samad, mis funktsioonil T.DIST.2T ($|t|$ ja $n-2$), kolmas argument on arv 2 (tähistab seda, et testime kahepoolset hüpoteesi $r \neq 0$, mitte seda, kas $r > 0$ või $r < 0$).

6. Tehke formaalne otsus, kumb püstitatud hüpoteesidest on õige ja miks.

A'la: p(Pikkus;Jalanr) 8,82306E-13 < 0,05 => H₁: tudengite pikkus ja jalanumber on seotud

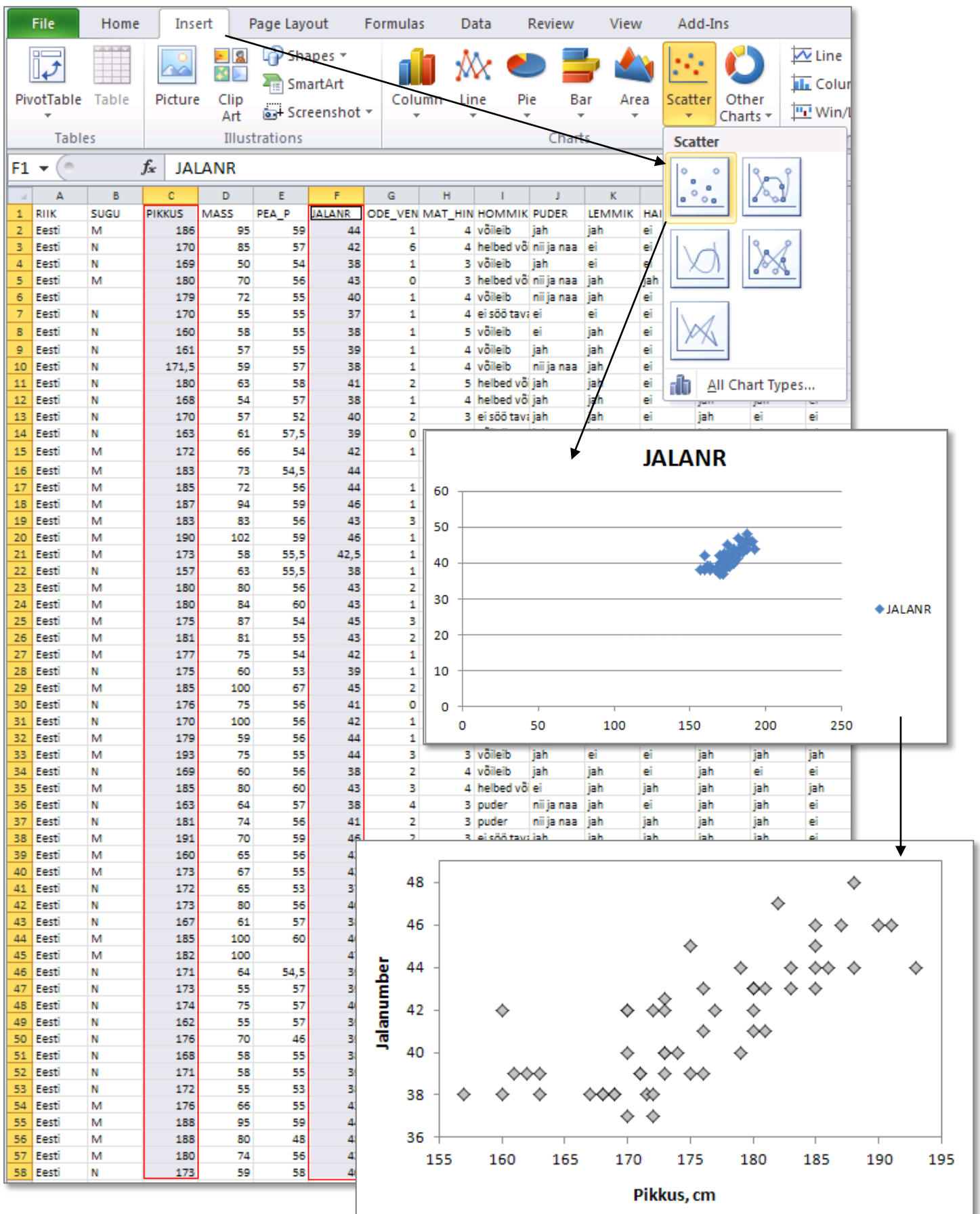
Märkus. 8,82306E-13 = 8,82... × 10⁻¹³

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on tugev positiivne statistiliselt oluline seos ($r = 0,780$; $p < 0,001$).

Ülesande 2 tööjuhend

Illustreerige tunnuste 'Pikkus' ja 'Jalanr' vahelist seost hajuvus- ehk punktdiagrammiga.



Ülesande 3 tööjuhend

1. Kasutades statistikaprotseduuri *Correlation* (Data-sakk -> Data analysis... -> Correlation) leidke korraga kõigi andmestikku kuuluvate pidevate arvturnuste (pikkus -- jalanumber) vahelised korrelatsioonikordajad.

The screenshot shows the Excel interface with the Data Analysis toolpak installed. The 'Data Analysis' dialog box is open, and 'Correlation' is selected. The 'Correlation' dialog box is also open, showing the 'Input Range' as '\$C\$1:\$F\$58' and 'Labels in first row' checked. The 'Output Range' is set to '\$U\$38'. A small table at the bottom shows the resulting correlation matrix.

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,636375299	1		
PEA_P	0,22408177	0,403565	1	
JALANR	0,779801717	0,752528	0,262844	1

Tulemus:

2. a) Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?
 b) Kas jalanumber on tugevamini seotud pikkusega või kehamassiga?
 c) Millise tunnusega on enim seotud peaümbermõõt?

Kirjeldage mõnda seost (pange vastavad laused ka kirja)!