

## 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 (inglisekeeles *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.

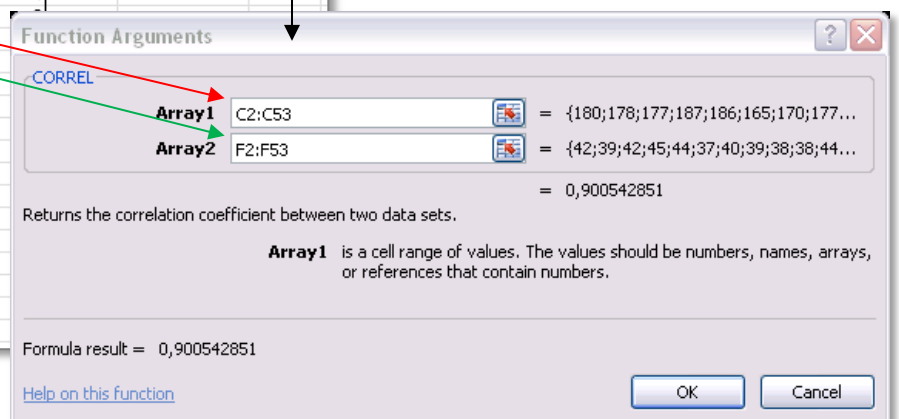
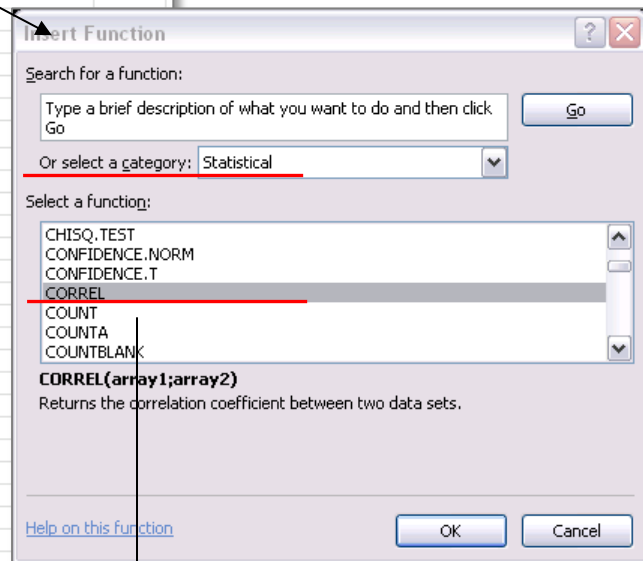
2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

- Kogenumad *MS Exceli* kasutajad võivad kohe sisestada arvutamiskäsu kujul  
=CORREL (C2 : C53 ; F2 : F53)

V1	A	B	C	D	E	F	G	T	U	V
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEI			r(Pikkus;Jalanr)
2	Eesti	N	180	78	56	42	2			
3	Eesti	N	178	65	56	39	4			
4	Eesti	M	177	70	57	42	3			
5	Eesti	M	187	75	45	45	3			
6	Eesti	M	186	74	43	44	1			
7	Eesti	N	165	62	42	37	3			
8	Eesti	N	170	67	55,5	40	4			
9	Eesti	N	177	59	42	39	1			
10	Eesti	N	166	47	55	38	2			
11	Eesti	N	165	55	42	38	1			
12	Eesti	M	180	68	51	44	0			
13	Eesti	N	161	49	56	37	2			
14	Eesti	N	168	54	50	40	0			
15	Eesti	N	167	67	56	40	2			
16	Eesti	N	160	50	55	38	0			
17	Eesti	N	164	53	59	38	1			
18	Eesti	M	194	105	57	47	4			
19	Eesti	M	178	65	53	42	3			
20	Eesti	M	177	90	57	44	1			
21	Eesti	N	171	57	50	38	2			
22	Eesti	M	187	99	58	45	1			
23	Eesti	M	189	81	54	46	1			
24	Eesti	M	186	98	56	45	0			
25	Eesti	N	171	65	55	38	1			
26	Eesti	M	183	110	57	43	0			
27	Eesti	M	193	100	58	46,5	1			
28	Eesti	N	183	79	54	42	4			
29	Eesti	N	177	75	55	40	1			
30	Eesti	N	164	54	52	37	1			
31	Eesti	N	170	60	55	39	2			
32	Eesti	N	175	65	57	38	2			
33	Eesti	N	168	69	55	39	1			
34	Eesti	M	186	94	58	46	1			
35	Eesti	N	160	55	55	36	2			
36	Eesti	N	175	69	53	39	8			
37	Eesti	N	151	53	52	36	3			
38	Eesti	M	198	110	60	47	4			
39	Eesti	M	174	120	56	45				
40	Eesti	N	180	80	58	42				
41	Eesti	N	171	72		41				
42	Eesti	N	169	80	57	40				
43	Eesti	N	160	64	54	38				
44	Eesti	N	170	65	55	39				
45	Eesti	N	176	58	54	40				
46	Eesti	N	167	55	54	39				
47	Eesti	N	160	59	56	38				
48	Eesti	N	172	80	54	39				
49	Eesti	N	170	60		39				
50	Eesti	N	176	57	56	41				
51	Eesti	N	168	62	56	38,5				
52	muu	M	189	78	57	47				
53	Eesti	N	173	68	45	40				

=CORREL (C2 : C53 ; F2 : F53)

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



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)	0,900543
Pikkuse ja jalanumbri vahel on ..... seos.	
St, et mida suurem on pikkus, seda ..... on keskmiselt ka jalanumber.	
Hüpoteeside paar	
H <sub>0</sub> : ..... (ehk matemaatiliselt r = 0)	
H <sub>1</sub> : ..... (ehk matemaatiliselt r ≠ 0)	

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

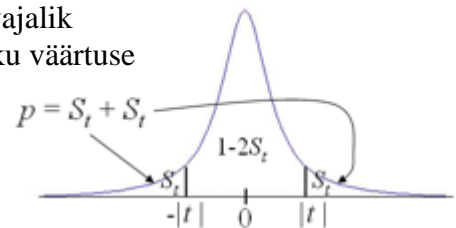
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

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<sub>t</sub> 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)	0,900543
Pikkuse ja jalanumbri vahel on tugev positiivne seos.	
St, et mida suurem on pikkus, seda suurem on keskmiselt ka	
Hüpoteeside paar	
H <sub>0</sub> : Pikkus ja jalanumber ei ole seotud (ehk matemaatiliselt r = 0)	
H <sub>1</sub> : 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)	0,900543			
2						
3		Pikkuse ja jalanumbri vahel on tugev positiivne seos.				
4		St, et mida suurem on pikkus, seda suurem on keskmiselt l				
5						
6		Hüpoteeside paar				
7		H <sub>0</sub> : Pikkus ja jalanumber ei ole seotud (ehk matemaatilisel				
8		H <sub>1</sub> : Pikkus ja jalanumber on seotud (ehk matemaatilisel r				
9						
10						
11		n(Pikkus;Jalanr)	52			
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)	0,900543								
2											
3		Pikkuse ja jalanumbri vahel on									
4		St, et mida suurem on pikkus, s									
5											
6		Hüpoteeside paar									
7		H <sub>0</sub> : Pikkus ja jalanumber ei ole									
8		H <sub>1</sub> : Pikkus ja jalanumber on sec									
9											
10											
11		n(Pikkus;Jalanr)	52								
12											
13		t(Pikkus;Jalanr)	14,64646								
14											
15		p(Pikkus;Jalanr)	=T.DIST.2T(V13;V11-2)								

Function Arguments

T.DIST.2T

X: V13 = 14,64645592

Deg\_freedom: V11-2 = 50

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

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

Formula result = 1,01554E-19

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 testimise 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)	1,01554E-19	< 0,05 => H <sub>1</sub> : tudengite pikkus ja jalanumber on seotud
------------------	-------------	---

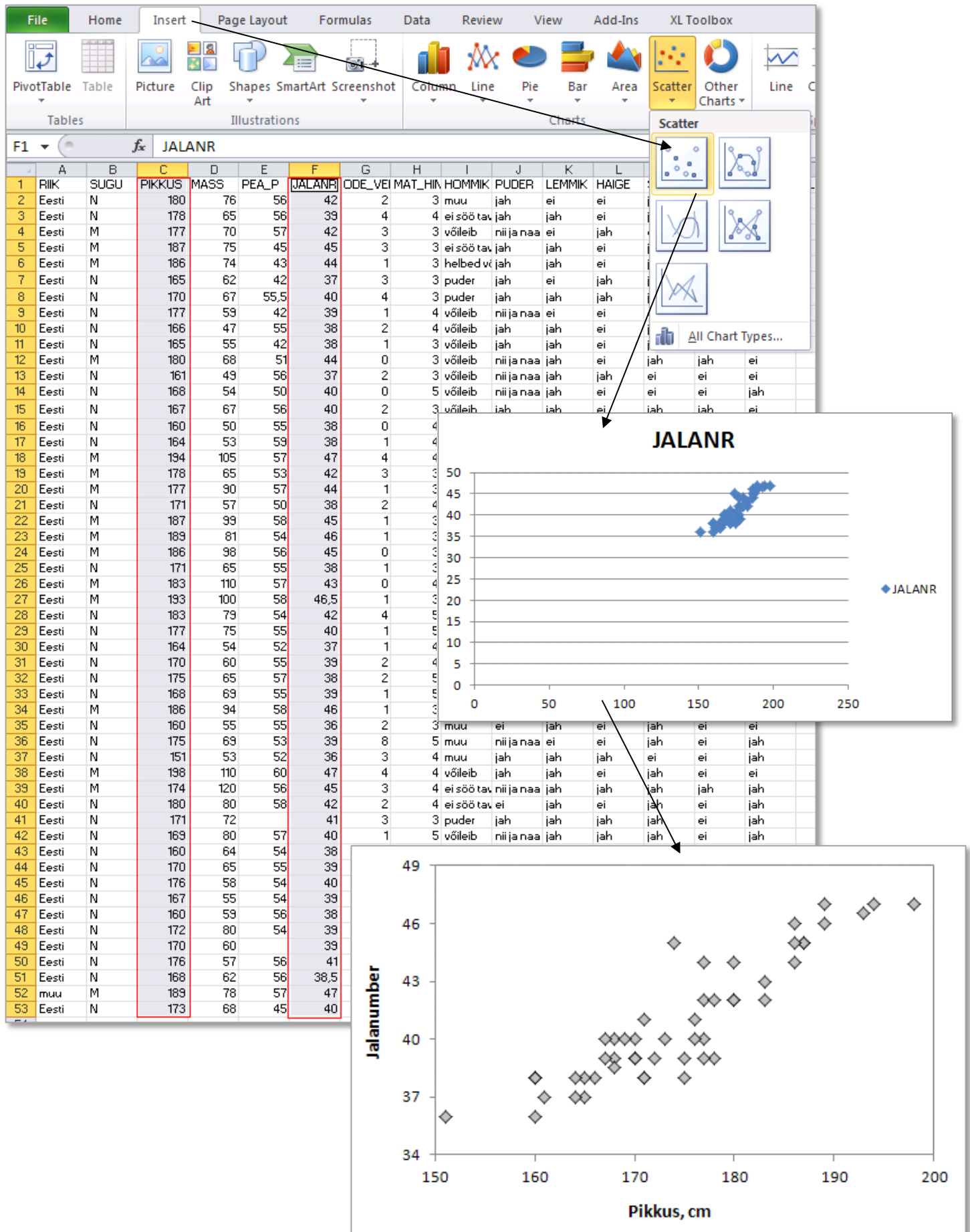
Märkus.  $1,01554E-19 = 1,01554... \times 10^{-19}$

## 7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on **tugev positiivne statistiliselt oluline** seos ( $r = 0,901$ ;  $p < 0,001$ ).

## 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 arv-tunnuste (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\$53' and the output range as '\$U\$39'. The 'Labels in first row' checkbox is checked. The spreadsheet data is visible in the background, with columns A through S and rows 1 through 53.

Tulemus:

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,738083	1		
PEA_P	0,1618981	0,347755	1	
JALANR	0,9005429	0,812778	0,235576	1

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 (nii seose suunda kui ka tugevust, pange vastavad laused kirja)!**