

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.

NB! Kuna „rebaseürituste“ tõttu jäi ära (lükkus edasi) tänase praktikumi teoreetilisi aluseid käsitlev loeng, on praktikumi ülesannete paremaks mõistmiseks kasulik üle vaadata 5. loengu (http://ph.emu.ee/~ktanel/VL_0781/VL1_loeng5.pdf) esimesed 7 slaidi (korrelatsioonanalüüs).

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.

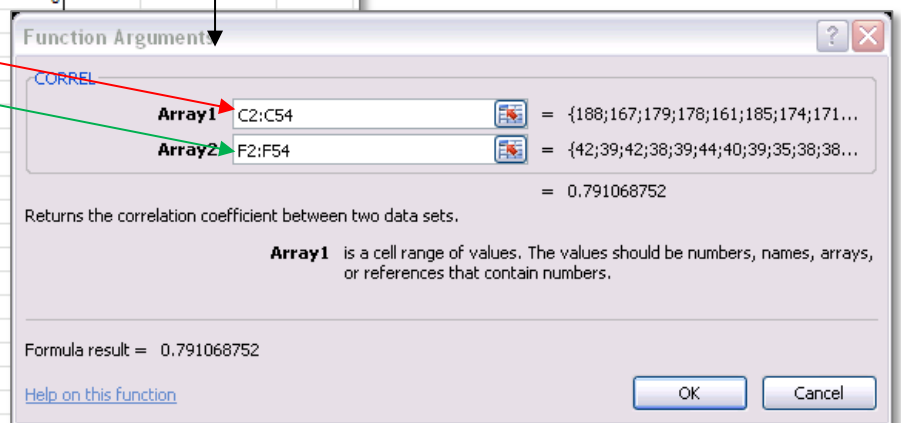
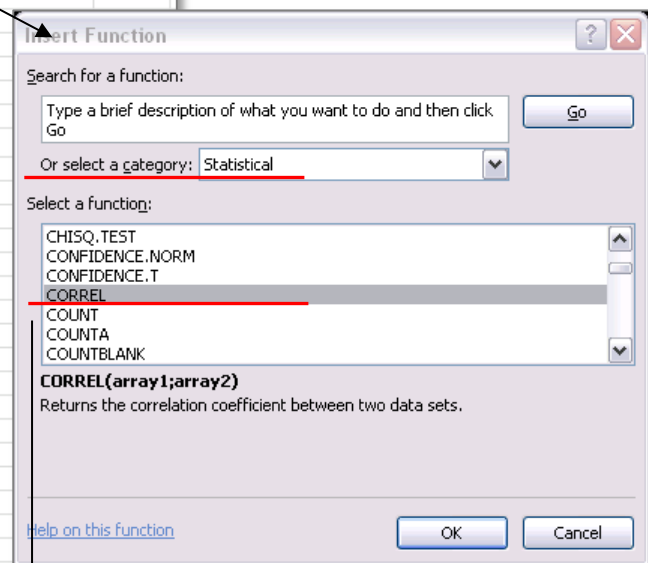
2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

- Kogenumad *MS Exceli* kasutajad võivad kohe sisestada arvutamiskäsu kujul

=CORREL (C2 : C54 ; F2 : F54)

- Vähemkogenumatel (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_VENI		r(Pikkus;Jalanr)	
2	Eesti	M	188	88	56	42	2			
3	Eesti	N	167	65	55	39	1			
4	Eesti	N	179	68	54	42	2			
5	Eesti	N	178	60	55	38	2			
6	Eesti	N	161	70	56.5	39	1			
7	Eesti	M	185	69	56	44	1			
8	Soome	N	174	58	58	40	10			
9	Eesti	N	171	55		39	2			
10	Eesti	N	165	55		35	0			
11	Soome	N	167	65	58.5	38	1			
12	Soome	N	163	69	58	38	2			
13	Soome	N	165	55	58	38	1			
14	Soome	N	169	61	58	36	2			
15	Soome	N	170	68	58	38	3			
16	Eesti	N	165	51	56	39	1			
17	Eesti	N	170	57	56	39	1			
18	Eesti	N	170	63	56	39	1			
19	Soome	N	162	65	58	39	1			
20	Eesti	N	168	80	56	37	2			
21	Eesti	N	177	60		39	2			
22	Eesti	N	168	73	53	39	2			
23	Eesti	N	165	50.1	52	38	3			
24	Eesti	M	188	80	58	44	2			
25	Eesti	N	162	63	56	40	1			
26	Soome	N	162	52	55	39	3			
27	Eesti	M	183	73	54	43	0			
28	Soome	N	161	47	62	37	0			
29			163	53	58	37	2			
30	Soome	N	179	80	62	43	3			
31	Soome	N	179	68	57	39	1			
32	Soome	N	176	80	55	40	1			
33	Soome	N	155	52	53	36	1			
34	Eesti	N	162	57	53	38	2			
35	Eesti	M	190	85	58	46	0			
36	Eesti	M	184	78	55	45				
37	Eesti	N	167	66	54	41				
38	Eesti	M	175	62	56.5	42				
39	Eesti	N	180	71	56	41				
40	Eesti	N	164	52	56	39				
41	Eesti	M	183	80	56	43				
42	Eesti	N	179	75	56	40				
43	Venemaa	N	167	46.5	55	37				
44	Eesti	N	172	64	55	40				
45	Soome	N	167	55	50	35				
46	Soome	N	171	75	56	39				
47	Soome	N	160	61	54	37				
48	Soome	N	168	75	58	40				
49	Soome	N	163	50	54	37				
50		N	169	65	57	39				
51	Soome	N	172	54	55	39				
52	Eesti	N	163	52	48	39				
53	Eesti	N	164	55	55.5	35				
54	Soome	N	156	48	50	37				



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)	0,791069
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.

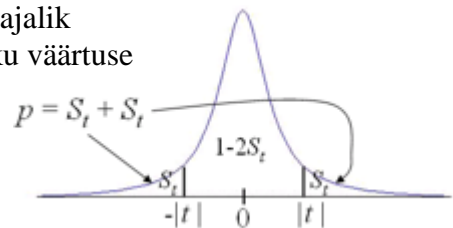
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

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)	0,791069
Pikkuse ja jalanumbri vahel on tugev positiivne seos.	
St, et mida suurem on pikkus, seda suurem on keskmiselt ka jalanumber.	
Hüpoteeside paar	
H ₀ : Pikkus ja jalanumber ei ole seotud (ehk matemaatiliselt r = 0)	
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)	0,791069			
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)	53			
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)	0,79106875								
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 seot									
9											
10											
11		n(Pikkus;Jalanr)	53								
12											
13		t(Pikkus;Jalanr)	9,23509762								
14											
15		p(Pikkus;Jalanr)	=T.DIST.2T(V13;V11-2)								

Function Arguments

T.DIST.2T

X: V13 = 9,235097616

Deg_freedom: V11-2 = 51

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

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

Formula result = 1,8092E-12

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) 1,8092E-12 < 0,05 => H₁: tudengite pikkus ja jalanumber on seotud

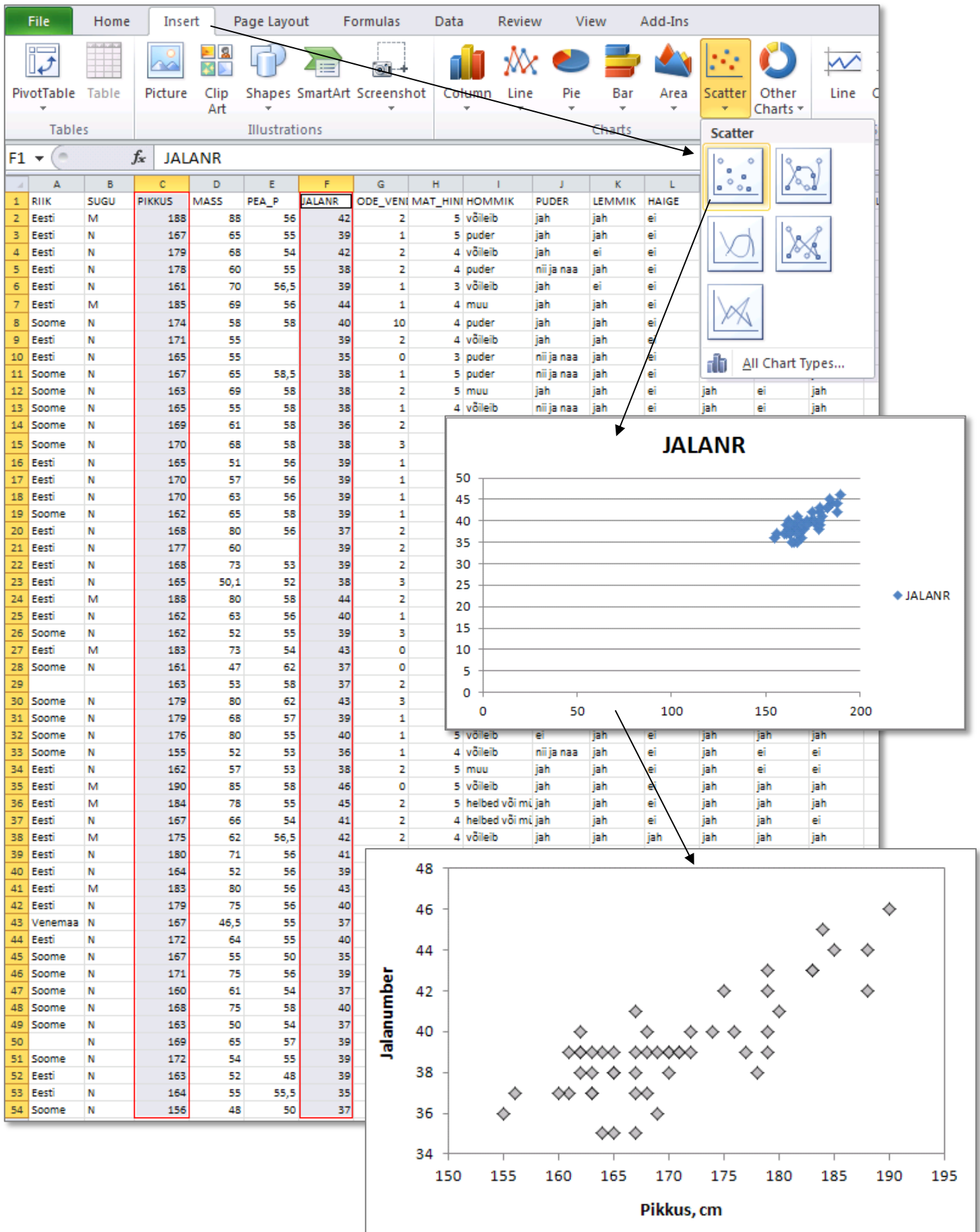
Märkus. 1,8092E-12 = 1,809... × 10⁻¹²

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on **tugev positiivne statistiliselt oluline** seos ($r = 0,791$; $p < 0,001$).

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 korruga 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 '\$C\$1:\$F\$54' and the output range '\$J\$39'. The 'Labels in first row' checkbox is checked. A red dashed box highlights the 'Labels in first row' checkbox, and a red arrow points from it to the first row of the data table below.

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,70573066	1		
PEA_P	0,23834289	0,303202	1	
JALANR	0,79106875	0,671693	0,215265	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 (nii seose suunda kui ka tugevust, pange vastavad laused kirja)!