

Biomeetria praks 5

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excel*'is 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 Excel*'i 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 Excel*'i 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 sooviks leida tudengite pikkuste ja jalanumbrite vahelist lineaarset korrelatsioonikordajat – *MS Excel*'i töölehele võiks siis trükkida


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

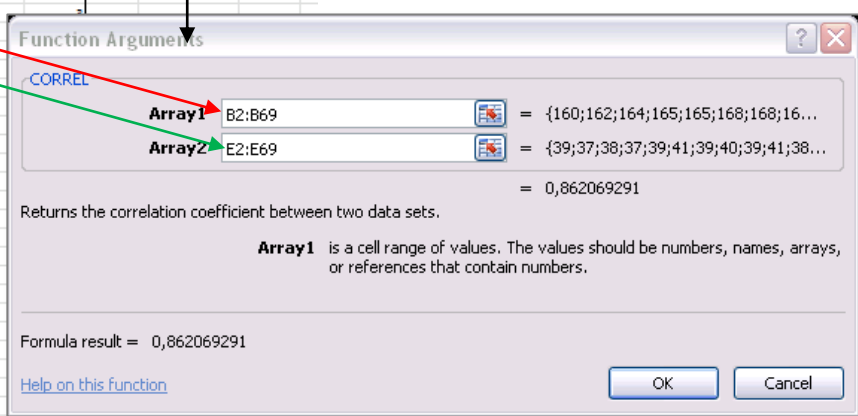
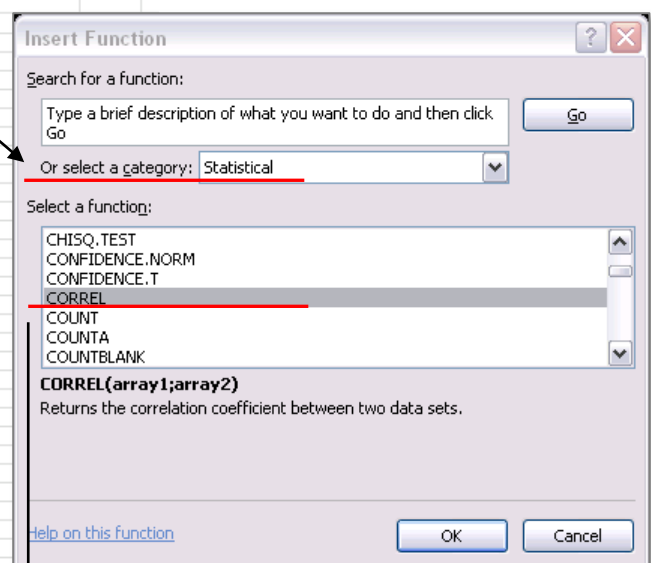
Seejärel pange kursor tühja lahtrisse, millesse soovite vastavat korrelatsioonikordajat arvutada.

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

Kogenumad *MS Excel*'i kasutajad võivad koheselt sisestada arvutamiskäsu kujul
 =CORREL (B2 : B69 ; E2 : E69)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	SUGU	PIKKUS	MASS	PEE_YMB	JALANR	ODE_VENI_MAT_HIN														r(Pikkus;Jalanr)
2	N	160	86	50	39	0-1														
3	N	162	80	55	37	0-1														
4	N	164	51	55	38	0-1														
5	N	165	54	54	37	0-1														
6	N	165	68	55	39	2-3														
7	N	168	58	56	41	>3														
8	N	168	68	52	39	2-3														
9	N	169	60	50	40	0-1														
10	N	169	65	55	39	2-3														
11	N	169	79	53	41	2-3														
12	N	170	62	58	38	0-1														
13	N	170	65	53	39	0-1														
14	N	170	70	55	40	2-3														
15	N	172	65	53	40	2-3														
16	M	173	78	48	44	>3														
17	M	173	80	57	44	0-1														
18	M	175	62	56	43	0-1														
19	M	175	74	58	43	2-3														
20	M	175	77	56	43	0-1														
21	M	178	110	58	42	0-1														
22	M	180	80	56	43	0-1														
23	M	181	88	56	43	0-1														
24	M	182	74	54	43	>3														
25	M	184	63	57	44	0-1														
26	M	192	89	55,5	46															
27	M	199	92	52	45	0-1														
28	M	195	83	56	46	>3														
29	N	153	47	53	37	0-1														
30	N	159	57	54	39	0-1														
31	N	160	52	53,5	38	2-3														
32	N	162	49	55	36	>3														
33	N	163	55	55	37	2-3														
34	N	165	70	52	39	>3														
35	N	166	56	54	38	>3														
36	N	166	60	54	38	2-3														
37	N	169	64	56	40	2-3														
38	N	169	65	58,5	38	2-3														
39	N	169	70	55	40	2-3														
40	N	169	73	56	38	0-1														
41	N	170	55	53	38	2-3														
42	N	170	56	55	38	0-1														
43	N	170	64	56	39	2-3														
44	N	170	78	57	42	2-3														
45	N	171	70	54	40	0-1														
46	N	172	55	54	38	>3														
47	N	172	75	55	39	2-3														
48	N	174	64	56	39	2-3														
49	N	175	80	52	41	0-1														
50	N	176	60	55	39	0-1														
51	M	178	61	54	43	0-1														
52	M	178	59	58	42	2-3														
53	N	179	72	57,5	41	2-3														
54	M	180	80	57	43	2-3														
55	M	183	150	63	46	0-1														
56	M	185	72	57	43	>3														
57	M	185	77	55	44	>3														
58	M	187	98	57	43	0-1														
59	M	190	85	58	46	2-3														
60	M	193	100	57,5	46	0-1														
61	M	194	88	54	47	0-1														
62	N	160	56	53	38	0-1														
63	N	171	57	54	37	2-3														
64	N	165	81	55	41	2-3														
65	N	167	60	52,5	39	0-1														
66	N	168	54	56	37	0-1														
67	N	169	50	54	39	0-1														
68	N	169	63	53	39	2-3														
69	N	172	58	57	38	0-1														

Vähem kogenumatel (eelnevast valemist 100%-liselt aru mittesaanutel) 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)	0,862069
Pikkuse ja jalanumbri vahel on seos.	
St, et mida suurem on pikkus, seda on keskmiselt jalanumber.	
Hüpoteeside paar	
H ₀ : (ehk matemaatiliselt)	
H ₁ : (ehk matemaatiliselt)	

4. Pange kirja

korrelatsioonikordaja kohta kontrollitav hüpoteeside paar (soovitavalt nii teksti kujul kui ka matemaatiliselt).

Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

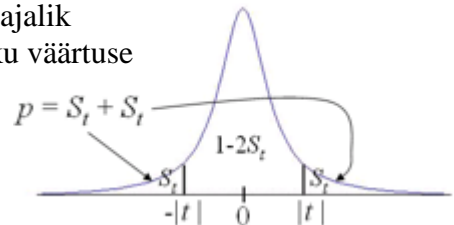
Korrelatsioonikordaja 0-st erinevuse testimiseks (st seose statistilise olulisuse testimiseks)

MS Excel³is 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)	0,862069
Pikkuse ja jalanumbri vahel on seos.	
St, et mida suurem on pikkus, seda on keskmiselt jalanumber.	
Hüpoteeside paar	
H ₀ : (ehk matemaatiliselt)	
H ₁ : (ehk matemaatiliselt)	
n(Pikkus;Jalanr)	
t(Pikkus;Jalanr)	
p(Pikkus;Jalanr)	

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

	R	S	T	U	V	W	X
1		r(Pikkus;Jalanr)	0,862069				
2							
3		Pikkuse ja jalanumbri vahel on seos.					
4		St, et mida suurem on pikkus, seda on keskmiselt jalanumber.					
5							
6		Hüpoteeside paar					
7		H ₀ : (ehk matemaatiliselt)					
8		H ₁ : (ehk matemaatiliselt)					
9							
10							
11		n(Pikkus;Jalanr)	68				
12							
13		t(Pikkus;Jalanr)	=ABS(T1*SQRT(T11-2)/SQRT(1-T1*T1))				
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 $(n - 2)$.

	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1		r(Pikkus;Jalanr)	0,862069										
2													
3		Pikkuse ja jalanumbri vahel on ...											
4		St, et mida suurem on pikkus, se											
5													
6		Hüpoteeside paar											
7		H ₀ : (ehk matemaatiliselt)											
8		H ₁ : (ehk matemaatiliselt)											
9													
10													
11		n(Pikkus;Jalanr)	68										
12													
13		t(Pikkus;Jalanr)	13,81929										
14													
15		p(Pikkus;Jalanr)	=T.DIST.2T(T13;T11-2)										

Function Arguments

T.DIST.2T

X: T13 = 13,81929025

Deg_freedom: T11-2 = 66

Formula result = 3,72799E-21

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

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

Help on this function

OK Cancel

NB! Vanemates Excel'i 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) 3,72799E-21 < 0,05 => H₁: tudengite pikkus ja jalanumber on seotud

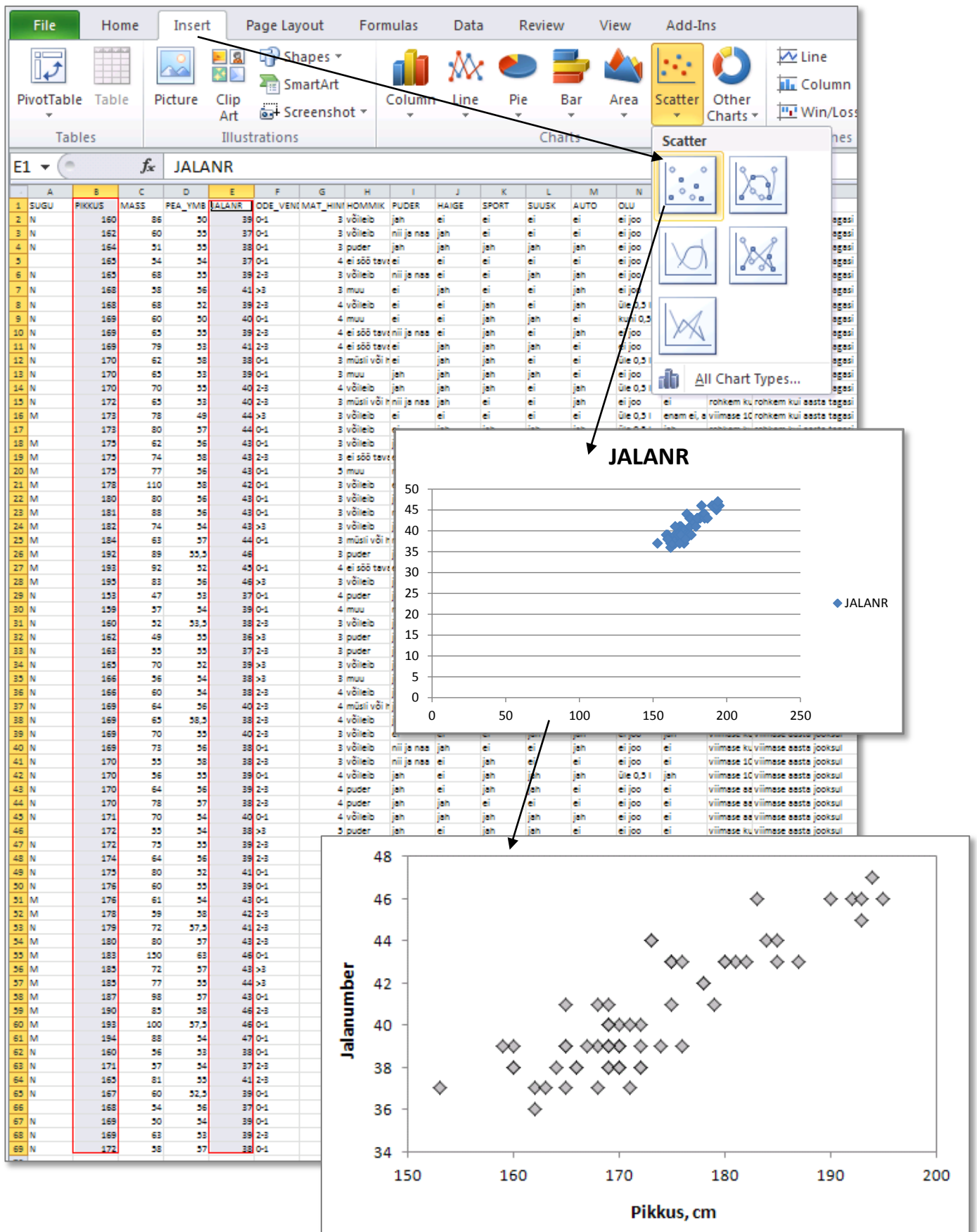
Märkus. 3,72799E-21 = 3,72... × 10⁻²¹

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on tugev positiivne statistiliselt oluline seos ($r = 0,862$; $p < 0,001$).

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' task pane open. The 'Correlation' tool is selected. The 'Input Range' is set to '\$B\$1:\$E\$69'. The 'Grouped By' option is set to 'Columns'. The 'Labels in first row' checkbox is checked. The 'Output Range' is set to '\$S\$18'. A scatter plot of height (cm) vs. number of steps is visible on the right side of the spreadsheet.

Tulemus:

	PIKKUS	MASS	PEA_YMB	JALANR
PIKKUS	1			
MASS	0,615526361	1		
PEA_YMB	0,360799149	0,366404	1	
JALANR	0,862069291	0,705013	0,258405	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 (pange vastavad laused ka kirja)!