

## 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 lõppjärgeldus.

#### 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 arv-tunnuste (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'.

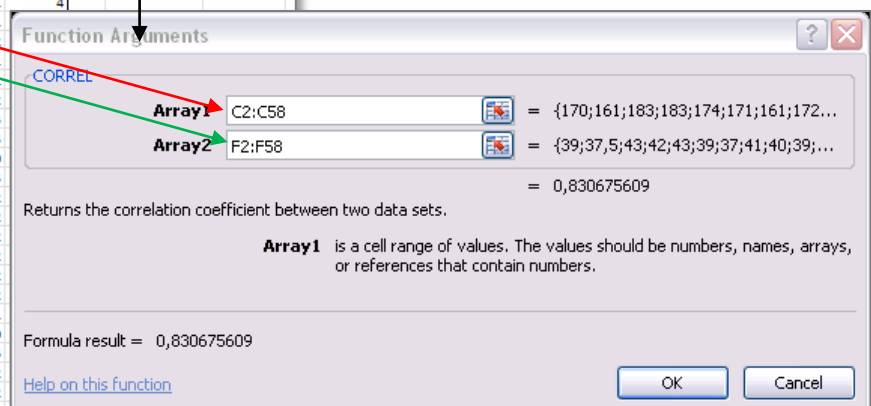
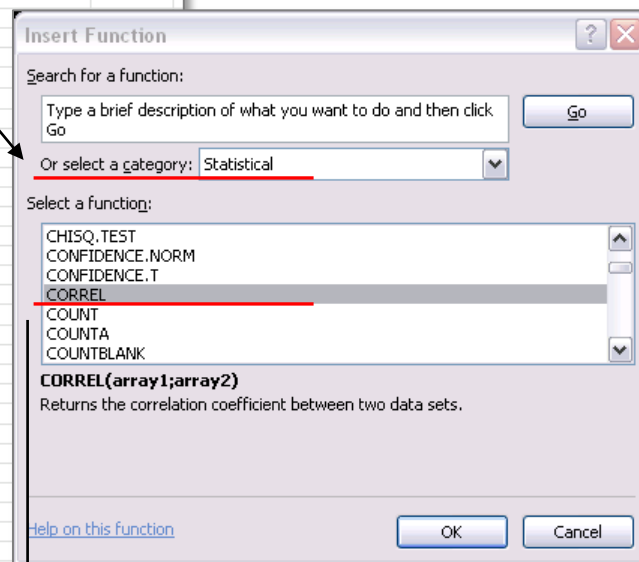
Seejärel pange kursor tühja lahtrisse, millesse soovite vastavat korrelatsioonikordajat arvutada.

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 kohealt sisestada arvutamiskäsu kujul  
=CORREL (C2 : C58 ; F2 : F58)

	A	B	C	D	E	G	H	T	V
	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEN	MAT_HIN	r(Pikkus;Jalanr)
2	Eesti	N	170	63	57	39	1	5	
3	Eesti	N	161	57	59	37,5	2	4	
4	Eesti	M	183	80	48	43	2	3	
5	Eesti	N	183	81	58	42	2	5	
6	Eesti	M	174	74	56	43	0	4	
7	Eesti	N	171	70	57	39	3	4	
8	Eesti	N	161	59	55	37	1	5	
9	Eesti	N	172	80	56	41	2	4	
10	Eesti	N	183,5	68	55,5	40	3	2	
11	Eesti	N	171	57	56	39	0	4	
12	Soome	N	175	68	56	40	2	4	
13	Eesti	N	168	70	57	40	2	5	
14	Eesti	N	152	50	52	37	1	3	
15	Eesti	N	164	55	54	37	2	4	
16	Eesti	N	160	60	55	37	2	3	
17	Eesti	N	163,3	67	56	39	3	5	
18	Eesti	N	163	53	56	37	1	4	
19	Soome	N	162	62	55	37	2	1	
20	Soome	N	158	65	56	37	4	3	
21	Eesti	N	170	57	57	39	8	4	
22	Eesti	N	170	58	56	37,5	2	3	
23	Eesti	N	162	57	52	38	3	4	
24	Eesti	N	169	63	56	39	1	4	
25	Eesti	N	176	67	54	41	0	5	
26	Eesti	N	173	64	54,5	41	1	5	
27	Eesti	N	170	63	54	38	2	4	
28	Eesti	N	167	73	56	39	1	3	
29	Eesti	N	167	58	56	39	2	4	
30	Eesti	M	178	80	59	42	1	4	
31	Eesti	M	189	71	60	47	2	3	
32	Eesti	M	181	61	55	46	1	4	
33	Eesti	M	189	71	60	47	2	3	
34	N	166	64	57	39	1	4		
35	Soome	N	169	65	57	39	1	4	
36	Soome	N	180	65	58	38	3	5	
37	Soome	N	185	59	59	41	4	4	
38	Soome	N	175	60	56	39	1	4	
39	Eesti	N	178	77	57	39	1	1	
40	Eesti	M	175	64	52	39	2	2	
41	Eesti	M	178	65	57	42	1	1	
42	Eesti	N	168	52	50	37,5	1	1	
43	N	160	62	53	36	2			
44	Eesti	M	185	80	57	45	3		
45	Eesti	N	175	85	59	41	3		
46	Eesti	N	167	65	55	39	0		
47	Eesti	N	176	74	57	43	3		
48	Eesti	N	161	62	56	38	2		
49	Soome	N	163	52	55	36	2		
50	Eesti	N	161	62	56	38	2		
51	Eesti	N	161,5	55	54	38	2		
52	Eesti	N	163	53	53	37	2		
53	Eesti	N	172	65	56	39	2		
54	Eesti	M	184	83	58	44	1		
55	Eesti	N	164	75	60	39	0		
56	Eesti	N	168	57	56	36	5		
57	Eesti	N	168	60	56	38	2		
58	Eesti	N	160	49	52	37	2		

Vähem kogenumatel (eelnevast valemist 100%-liselt aru mittesaanutel) on soovitatav klikkida nupul  ja jätkata vastavalt järgnevale 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“.

$r(\text{Pikkus};\text{Jalanr})$	0,830676
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).

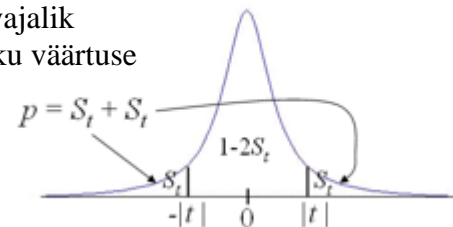
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

Korrelatsioonikordaja 0-st erinevuse testimiseks (st seose statistilise olulisuse testimiseks) MS 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 ' $n(\text{Pikkus};\text{Jalanr})$ ' 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).

$r(\text{Pikkus};\text{Jalanr})$	0,830676
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 .....)	
$n(\text{Pikkus};\text{Jalanr})$	
$t(\text{Pikkus};\text{Jalanr})$	
$p(\text{Pikkus};\text{Jalanr})$	

b) Lahtri  $t$ (Pikkus;Jalanr) järel tuleb sisestada valem teststatistiku absoluutväärtuse arvutamiseks:

	T	U	V	W	X	Y	Z
1		r(Pikkus;Jalanr)	0,830676				
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 <sub>0</sub> : ..... (ehk matemaatiliselt .....)					
8		H <sub>1</sub> : ..... (ehk matemaatiliselt .....)					
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) järel 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	AE	AF
1		r(Pikkus;Jalanr)	0,830675609										
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 <sub>0</sub> : ..... (ehk matemaatiliselt .....)											
8		H <sub>1</sub> : ..... (ehk matemaatiliselt .....)											
9													
10													
11		n(Pikkus;Jalanr)	57										
12													
13		t(Pikkus;Jalanr)	11,0648962										
14													
15		p(Pikkus;Jalanr)	=T.DIST.2T(V13;V11-2)										

**Function Arguments**

T.DIST.2T

X: V13 = 11,0648962

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 = 1,31399E-15

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:  $p(\text{Pikkus;Jalanr}) \ 1,31399\text{E-}15 < 0,05 \Rightarrow H_1: \text{tudengite pikkus ja jalanumber on seotud}$

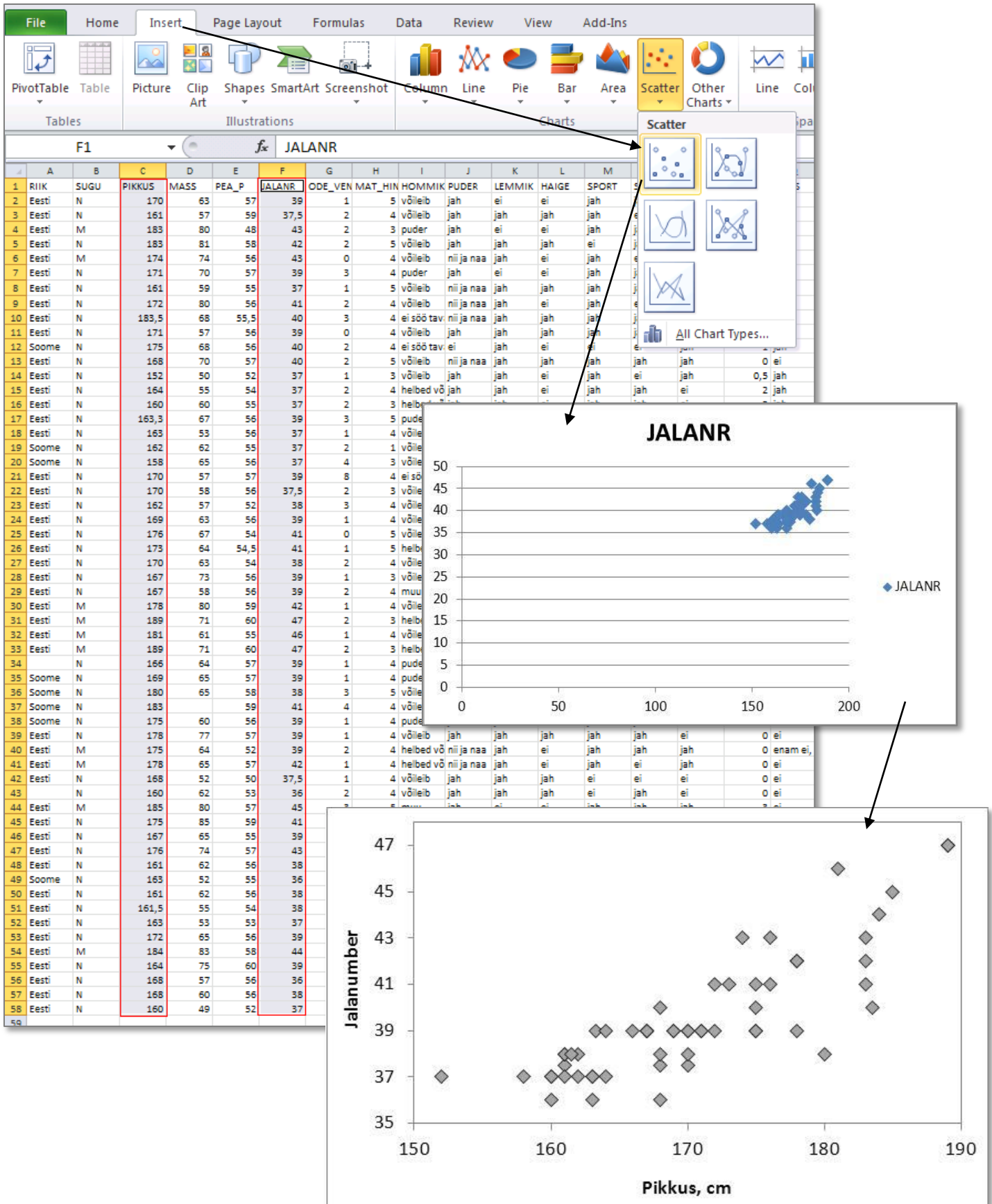
← Märkus.  $1,31399\text{E-}15 = 1,31... \times 10^{-15}$

**7. Pange kirja päris lõppjärelendus.**

A'la: tudengite pikkuse ja jalanumbri vahel on statistiliselt oluline tugev positiivne seos ( $r = 0,831; p < 0,001$ ).

## 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 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 under the 'Analysis Tools' list. The 'Correlation' dialog box is also open, showing the 'Input Range' as '\$C\$1:\$F\$58' and the 'Output Range' as '\$U\$20'. The 'Labels in first row' checkbox is checked. A scatter plot is visible in the background, showing a positive correlation between 'Pikkus, cm' (Height) and 'Jalanumber' (Shoe size).

Tulemus:

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,648811715	1		
PEA_P	0,360504788	0,439157	1	
JALANR	0,830675609	0,653976	0,368109	1

2. Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?

Kas jalanumber on tugevamini seotud pikkusega või kehamassiga?

Millise tunnusega on enim seotud peaümberrõõd?

**Kirjeldage mõnda seost (pange vastavad laused ka kirja)!**