

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 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 sooviks leida tudengite pikkuste ja jalanumbrite vahelist lineaarset korrelatsioonikordajat – *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.

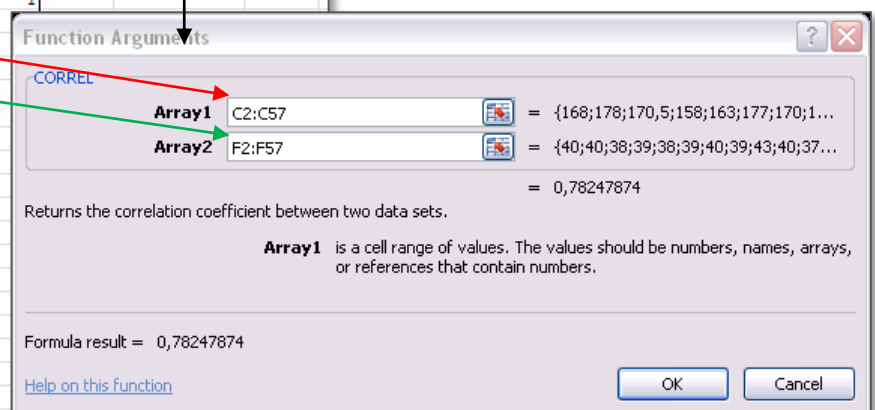
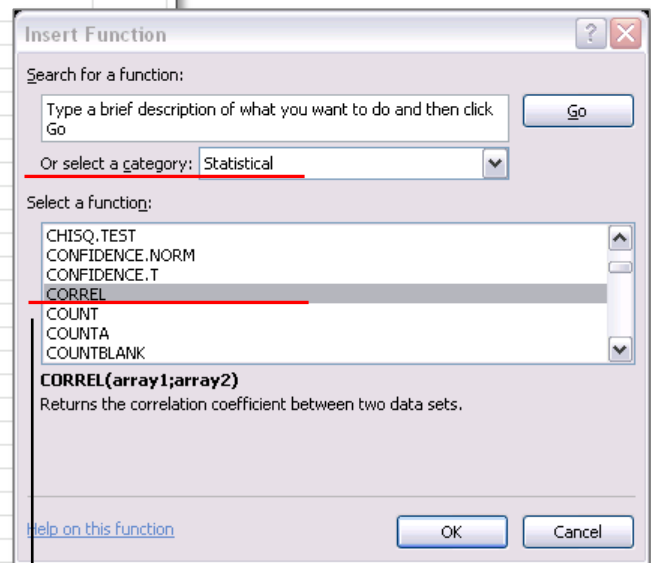
2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

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

=CORREL (C2 : C57 ; F2 : F57)

- 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	N	168	62	53	40	1			
3	Eesti	N	178	75	56	40	4			
4	Eesti	N	170,5	64	54	38	2			
5	Eesti	N	158	56	50	39	2			
6	Eesti	N	163	60	54	38	1			
7	Eesti	N	177	70	53	39	3			
8	Eesti	N	170	51	52	40	0			
9	Eesti	N	170	56		39	2			
10	Soome	M	183	90	60	43	2			
11	Soome	N	167	60	54	40	1			
12	Soome	N	157	46	54	37	2			
13	Soome	N	175	65	56	40	2			
14	Soome	N	157	70	56	36	2			
15	Soome	N	173	62	56	39	1			
16	Soome	N	161	65	58	39	3			
17	Eesti	N	169	68	55,5	39	2			
18	Eesti	N	178	56	56	40	1			
19	Eesti	N	164	62	55	37,5	1			
20	Eesti	N	159	55	54	37	2			
21	Eesti	N	164	55	54,5	37	1			
22	Eesti	N	174	74	58	40	3			
23	Eesti	N	171	67	66	38	1			
24	Soome	N	160	56	57	38	2			
25	Soome	N	170	76	56	39	2			
26	Soome	M	187	75	55	44	2			
27	Soome	N	165	55	56	37	4			
28	Eesti	N	168	55	57	39	2			
29	Eesti	N	159	48	52	34	1			
30	Soome	N	166	58	57	38	2			
31	Soome	N	169	59	58	38	1			
32	Soome	N	160	57	58	37	1			
33	Soome	N	178	70	53	41	2			
34	Soome	N	169	68	58	40	1			
35	Eesti	N	174	62	58	40	1			
36	Eesti	N	165	59	55	37	1			
37	Eesti	N	177	75	55	40				
38	Soome	N	170	80	59	39				
39	Eesti	N	173	75	55	40				
40	Eesti	N	175,5	66	55	39				
41	Eesti	N	172	66	56	40				
42	Eesti	N	165	60	57	39				
43	Eesti	N	170	60	58	38				
44	Eesti	N	168	55	58	39				
45	Soome	N	158	75	58	38				
46	Eesti	N	170	54	53	39				
47	Soome	N	181	72	57	42				
48	Soome	N	167	70	51,5	40				
49	Soome	N	166	49	55	38				
50	Soome	N	174	51	55	38				
51	Soome	N	160	60	57	38				
52	Soome	N	170	65	59	39				
53	Soome	N	170	66	58	38				
54	Eesti	M	171	75	56	39				
55	Eesti	M	177	69	55	43				
56	Eesti	N	168	65	55	39				
57	Eesti	N	172	62	53	39				



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)	0,782479
Pikkuse ja jalanumbri vahel on seos.	
St, et mida suurem on pikkus, seda on keskmiselt 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.

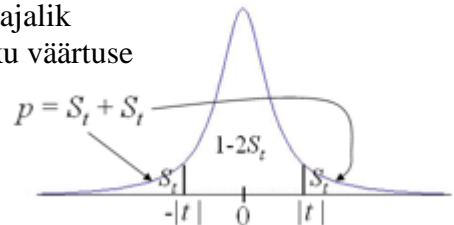
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

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 arvutatud (ehk nende tudengite arv, kelle kohta on teada nii nende pikkus kui ka jalanumber).

r(Pikkus;Jalanr)	0,782479
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)	0,782479			
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)	56			
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)	0,782479								
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)	56								
12											
13		t(Pikkus;Jalanr)	9,234351								
14											
15		p(Pikkus;Jalanr)	=T.DIST.2T(V13;V11-2)								

Function Arguments

T.DIST.2T

X: V13 = 9,234351301

Deg_freedom: V11-2 = 54

Formula result = 1,06091E-12

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

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

Help on this function

OK

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: $p(\text{Pikkus;Jalanr}) = 1,06091\text{E-}12 < 0,05 \Rightarrow H_1: \text{tudengite pikkus ja jalanumber on seotud}$

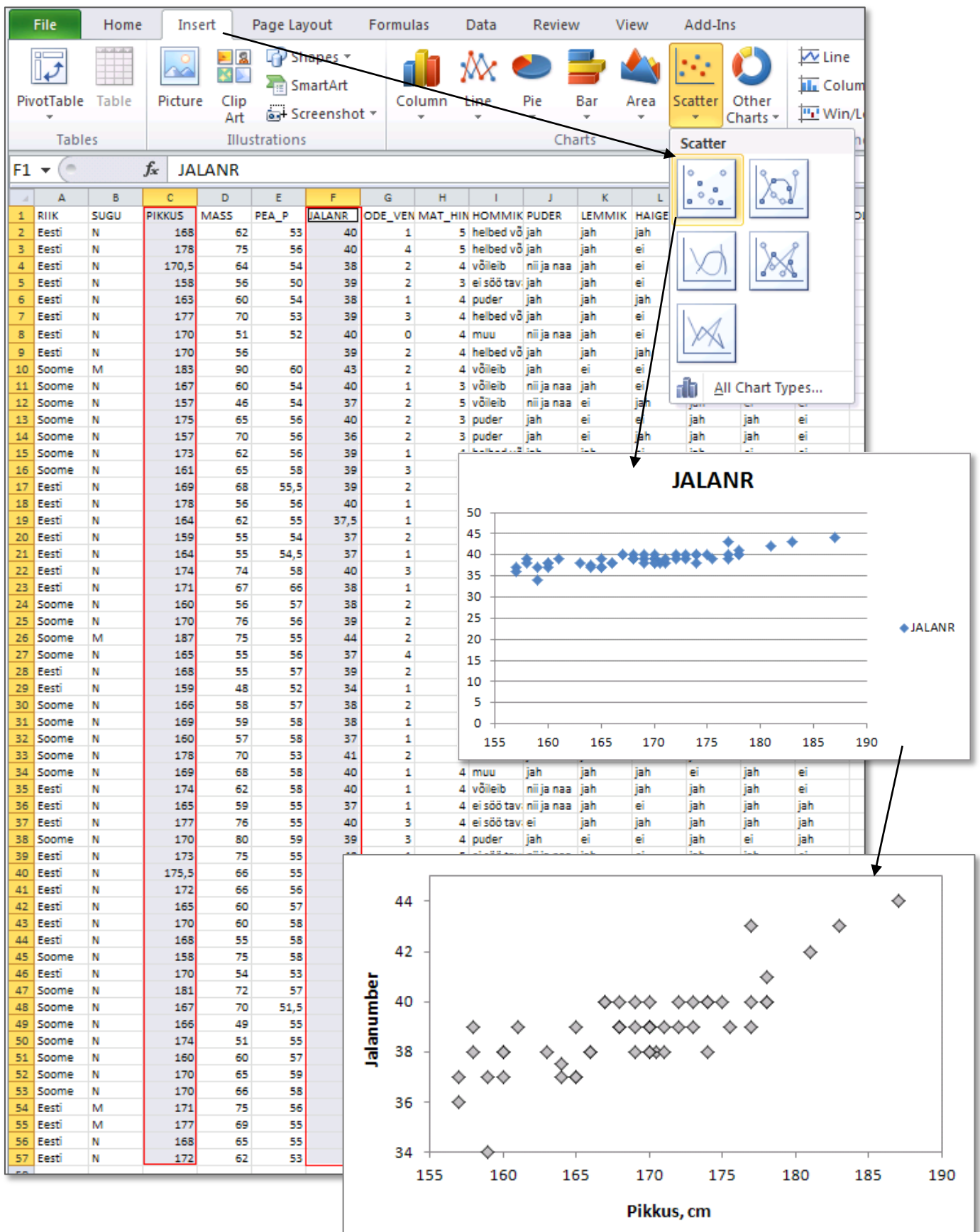
Märkus. $1,06091\text{E-}12 = 1,06\dots \times 10^{-12}$

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on **tugev positiivne statistiliselt oluline** seos ($r = 0,782$; $p < 0,001$).

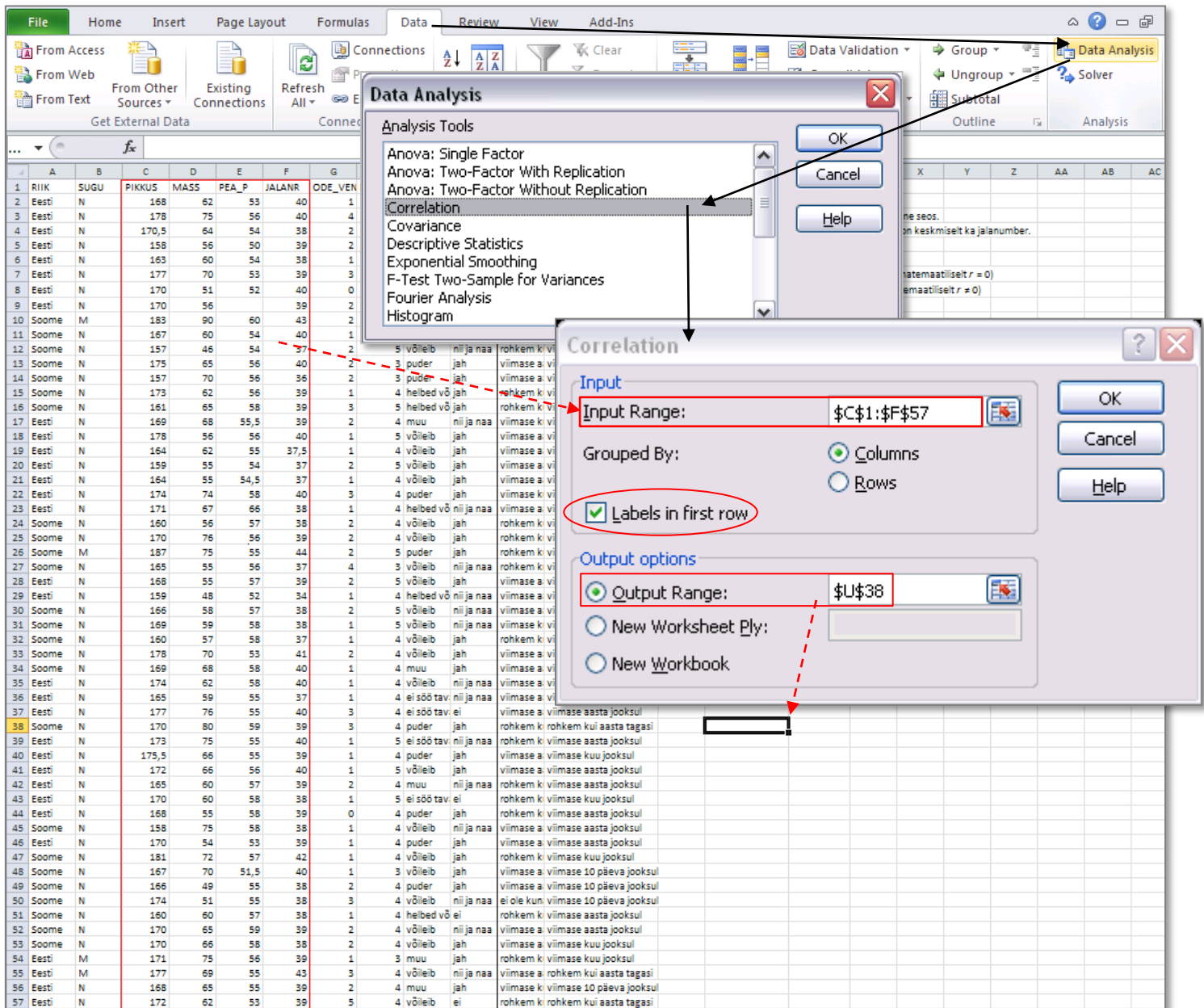
Ülesande 2 tööjuhend

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.



Tulemus:

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,523920934	1		
PEA_P	0,12246397	0,311647	1	
JALANR	0,78247874	0,565154	0,041412	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)!