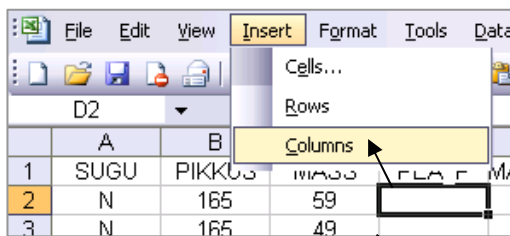


Biomeetria praks 5

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav (**elmites praktikumides puhastatud**) andmestik,
2. lisage uus tööleht (*Insert / Lisa -> Worksheet / Arvutustabel*), nimetage see ümber leheküljeks 'Praks5' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks5' ülemisse vasakusse nurka.
4. Lisage kehamassi ja peaümbermõõdu vahele uus veerg ja arvutage sinna (**analooqselt praktikumile 3**) kõigi tudengite kehamassiindeksid valemist $KMI = \text{Kehamass, kg} / (\text{Pikkus, m})^2$.
Kustutage veateadet #DIV/0! sisaldavate või 0-lise väärtusega lahtrite sisu.



	A	B	C	D	E	F
	SUGU	PIKKUS	MASS	KMI	PEA_P	MAT_HINN
1	N	165	59	=C2/((B2/100)^2)		4
2	N	165	49	17,9982		4
3	N	171	70	23,939	56	3
4	M	186	75	21,6788		4
5	M	182	73	22,0384	56	4
6	N	172,5	60	20,1638		4
7	N	167	75	26,8923		4
8	N	175	64	20,898	50	3
9	N	169	72	25,2092	55	3
10	M	174	93	30,7174	58	3
11	M	191	88	24,1221	60	3
12	M	175	70	22,8571	56	5
13	N	157	57	23,1247		5
14	M	187	70	20,0177	56	3
15	M	187	56	16,0142		3
16	N	164	50	18,5901		4
17	M	185	80	23,3747	56	3
18	M	179	89	27,7769	56	4
19	N	177	79	25,2163	53	3
20	M	191	80	21,9292	60	4
21	N	162		0	54	3
22	M	188	78	22,0688		4
23	N	168	55	19,487	50	3
24	N	175	70	22,8571	56	5
25	M	186	80	23,1241	66	3
26	N	169	55	19,257		3
27	N	171	54	18,4672	53,2	3
28	N	178	80	25,2493	56	5
29	N	157	48	19,4734		4
30	N	180	67	20,679	61	5
31	M	190	85	23,5457		3
32	N	173	70	23,3887		4
33	M			#DIV/0!		5
34	M	175	70	22,8571		3

Ülesanne 1.

Kas tudengite kehamass ja kehamassiindeks on seotud? Uurige seost *MS Excel*'i funktsioonide abil.

- Leidke tunnuste 'Mass' ja 'KMI' vaheline lineaarne korrelatsioonikordaja;
- kirjeldage kehamassi ja kehamassiindeksi vahelist seost leitud kordaja alusel;
- testige seose statistilist olulisust:
 - sõnastage null- ja alternatiivne hüpotees (pange need töölehele kirja),
 - kontrollige nende kehtimist (leidke teststatistik t , vaatluspaaride arv n ja nende alusel olulisuse tõenäosus p),
 - pange kirja lõppjärelendus.

Ülesanne 2.

Kasutades statistikaprotseduuri *Correlation (Tools/Tööriistad -> Data analysis...)* leidke korruga kõigi andmestikku kuuluvate pidevate arvtunnuste vahelised lineaarsed korrelatsioonikordajad.

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

Kirjeldage mõnda seost (pange vastavad laused ka kirja).

Ülesanne 3.

Illustreerige tunnuste 'Pikkus' ja 'KMI' vahelist seost hajuvusdiagrammiga (punktdiagrammiga).

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 kehamasside ja kehamassiindeksite vahelist lineaarset korrelatsioonikordajat – *MS Excel*'i töölehele võikski siis trükkida

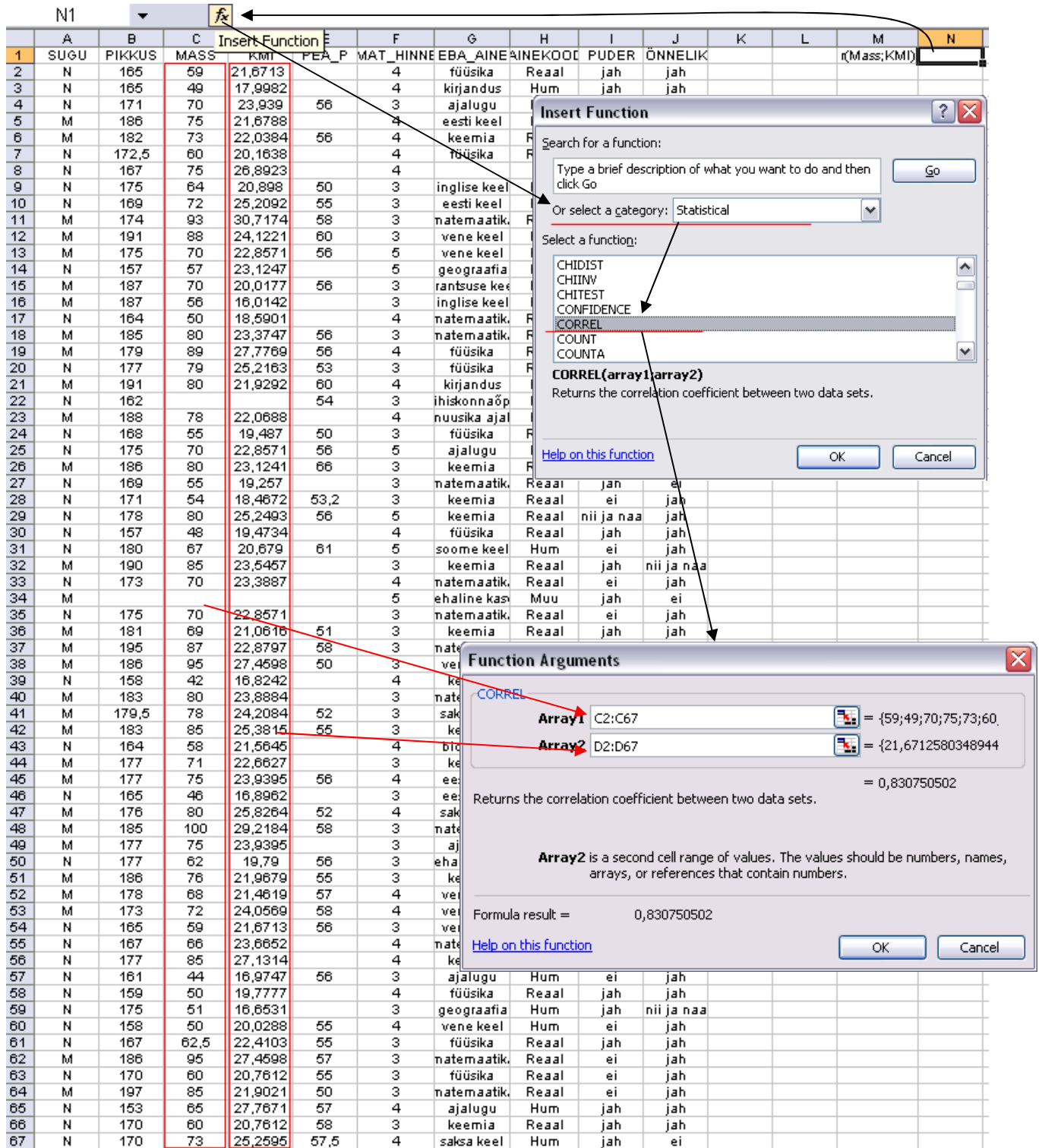
- '**Kehamassi ja kehamassiindeksi vaheline lineaarne korrelatsioonikordaja**'
- või lühemalt '**r(Mass;KMI)**' – lineaarset korrelatsioonikordajat tähistatakse enamasti tähega 'r'.

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 : C67 ; D2 : D67)

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



The screenshot shows an Excel spreadsheet with columns A through N. Column C contains mass values (e.g., 165, 171, 186) and column D contains height values (e.g., 172.5, 167, 175). Two dialog boxes are overlaid on the spreadsheet:

- Insert Function:** This dialog box is used to select the CORREL function. The 'Or select a category' dropdown is set to 'Statistical'. The 'Select a function' list shows 'CORREL' selected. A description below reads: 'CORREL(array1,array2) Returns the correlation coefficient between two data sets.'
- Function Arguments:** This dialog box shows the arguments for the CORREL function. 'Array1' is set to 'C2:C67' and 'Array2' is set to 'D2:D67'. The calculated result is shown as '= 0,830750502'. A description below reads: 'Returns the correlation coefficient between two data sets. Array2 is a second cell range of values. The values should be numbers, names, arrays, or references that contain numbers.'

3. Kirjeldage, milline on tudengite kehamasside ja kehamassiindeksi väärtuste vaheline seos

- kui tugev (nõrk / keskmise tugevusega / tugev),
- kas positiivne või negatiivne (mida see positiivne või negatiivne tähendab?).

r(Mass;KMI)	0,830751								
Kehamassi ja kehamassiindeksi vahel on seos.									
St, et mida suurem on kehamass, seda on keskmiselt kehamassiindeksi väärtus.									
Hüpoteeside paar									
H ₀ : (ehk matemaatiliselt)									
H ₁ : (ehk matemaatiliselt)									

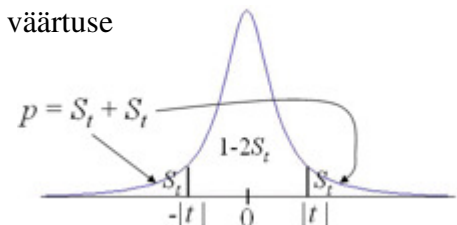
4. Pange kirja korrelatsioonikordaja kohta kontrollitav hüpoteeside paar (soovitavalt nii teksti kujul kui ka matemaatiliselt).

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 (so nende tudengite arv, kelle kohta on teada nii kehamassi kui ka kehamassiindeksi 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'is on p-väärtus leitav funktsiooniga TDIST (ABS (t) ; n-2 ; 2) .

5. Olulisuse tõenäosuse p leidmiseks vajalikke arvutusi on mugav teostada, kui kõik vajalikud suurused on töölehele kirja pandud.

Näiteks kujul:

r(Mass;KMI)	0,830751								
Kehamassi ja kehamassiindeksi vahel on seos.									
St, et mida suurem on kehamass, seda on keskmiselt kehamassiindeksi väärtus.									
Hüpoteeside paar									
H ₀ : (ehk matemaatiliselt)									
H ₁ : (ehk matemaatiliselt)									
n(Mass;KMI)									
t(Mass;KMI)									
p(Mass;KMI)									

- a) Lahtri 'n(Mass;KMI)' järele tuleks kirjutada nende tudengite arv, kelle andmete alusel on korrelatsiooni-kordaja arvatud (so nende tudengite arv, kelle kohta on teada nii kehamassi kui ka kehamassiindeksi väärtus – vt ise andmetest).

b) Lahtri 't(Mass;KMI)' järele tuleb sisestada valem teststatistiku absoluutväärtuse arvutamiseks:

	L	M	N	O	P	Q
1		r(Mass;KMI)	0,830751			
2						
3		Kehamassi ja kehamassiindeksi vahel on se				
4		St, et mida suurem on kehamass, seda on keskm				
5						
6		Hüpoteeside paar				
7		H ₀ : (ehk matemaatiliselt)				
8		H ₁ : (ehk matemaatiliselt)				
9						
10						
11		n(Mass;KMI)	64			
12						
13		t(Mass;KMI)	=ABS(N1*SQRT(N11-2)/SQRT(1-N1*N1))			

c) Lahtri 'p(Mass;KMI)' järele tuleb sisestada funktsioon TDIST kolme argumendiga:

- teststatistiku absoluutväärtus |t|,
- (vaatluspaaride arv) – 2, so vastava t-jaotuse parameeter (n – 2), ja
- arv 2 (mis näitab, et testime kahepoolset hüpoteesi – mida see tähendab?).

	L	M	N	O	P	Q	R	S	T	U	V
1		r(Mass;KMI)	0,830750502								
2											
3		Kehamassi ja kehamassiindeksi vahel									
4		St, et mida suurem on kehamass, sed									
5											
6		Hüpoteeside paar									
7		H ₀ : (ehk matemaatiliselt)									
8		H ₁ : (ehk matemaatiliselt)									
9											
10											
11		n(Mass;KMI)	64								
12											
13		t(Mass;KMI)	11,75136135								
14											
15		p(Mass;KMI)	=TDIST(N13;N11-2;2)								
16											
17											

Function Arguments

TDIST

X: N13 = 11,75136135

Deg_freedom: N11-2 = 62

Tails: 2 = 2

Result: = 2,02945E-17

Returns the Student's t-distribution.

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

Formula result = 2,02945E-17

[Help on this function](#)

OK Cancel

6. Tehke formaalne otsus, kumb püstitatud hüpoteesidest on õige ja miks.

A'la: p(Mass;KMI) 2,02945E-17 < 0,05 => H1: tudengite kehamass ja kehamassiindeks on seotud

7. Pange kirja päris lõppjärelus.

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

Ülesande 2 tööjuhend

1. Kasutades statistikaprotseduuri *Correlation (Tools/Tööriistad -> Data analysis... -> Correlation)* leidke korraga kõigi andmestikku kuuluvate pidevate arv-tunnuste vahelised korrelatsioonikordajad.

The screenshot shows an Excel spreadsheet with a data table and two dialog boxes. The data table has columns A-D with values for SUGU, PIKKUS, MASS, and KMI. The Correlation dialog box shows the input range as \$B\$1:\$E\$67, grouped by columns, with 'Labels in first row' checked, and the output range as \$M\$20.

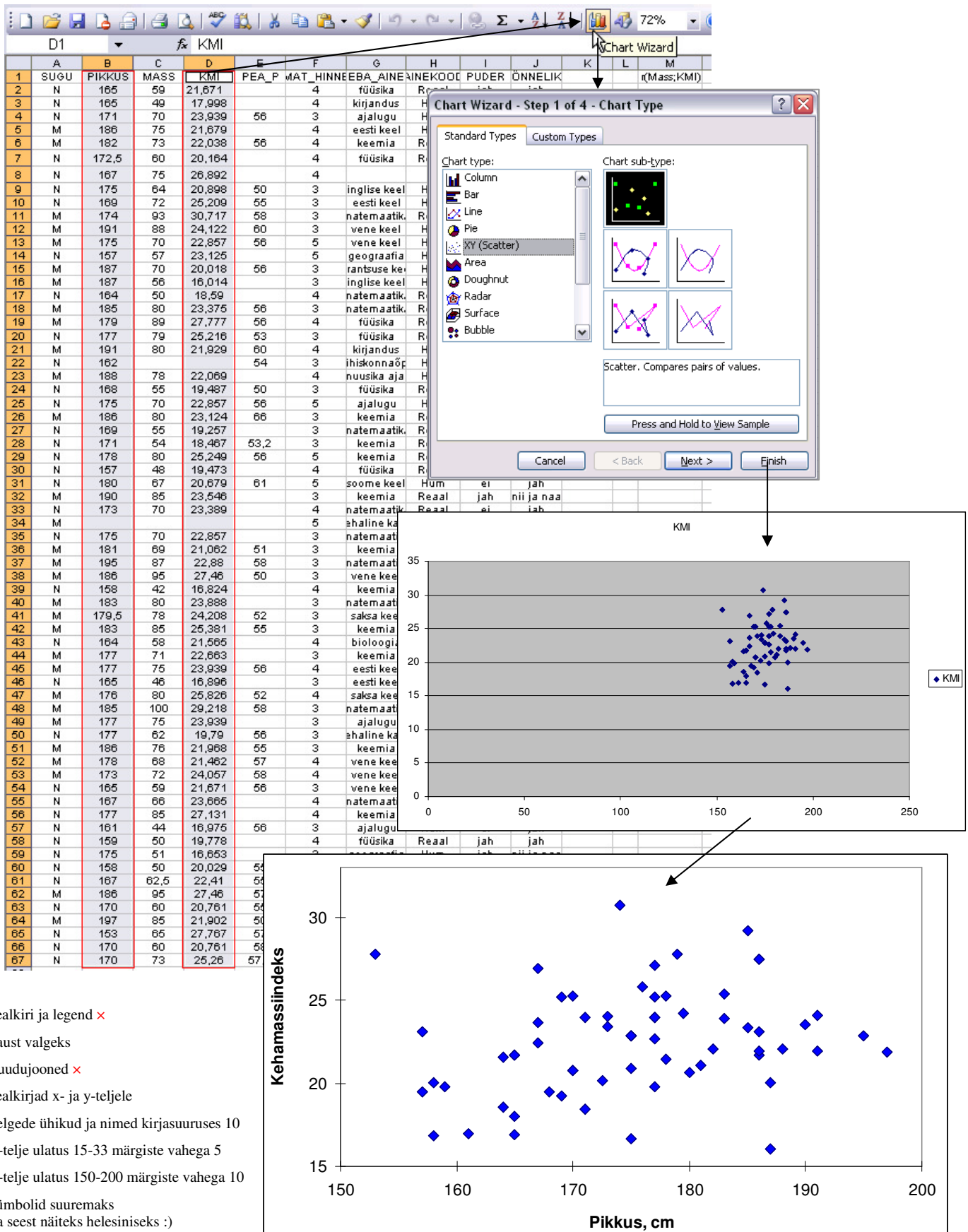
	A	B	C	D	PE
1	SUGU	PIKKUS	MASS	KMI	PE
2	N	165	59	21,671	
3	N	165	49	17,998	
4	N	171	70	23,939	
5	M	186	75	21,679	
6	M	182	73	22,038	
7	N	172,5	60	20,164	
8	N	167	75	26,892	
9	N	175	64	20,898	
10	N	169	72	25,209	
11	M	174	93	30,717	
12	M	191	88	24,122	
13	M	175	70	22,857	
14	N	157	57	23,125	
15	M	187	70	20,018	
16	M	187	56	16,014	
17	N	164	50	18,59	
18	M	185	80	23,375	
19	M	179	89	27,777	
20	N	177	79	25,216	
21	M	191	80	21,929	
22	N	162			
23	M	188	78	22,069	
24	N	168	55	19,487	
25	N	175	70	22,857	
26	M	186	80	23,124	
27	N	169	55	19,257	
28	N	171	54	18,467	5
29	N	178	80	25,249	
30	N	157	48	19,473	
31	N	180	67	20,679	
32	M	190	85	23,546	
33	N	173	70	23,389	
34	M				
35	N	175	70	22,857	
36	M	181	69	21,062	51
37	M	195	87	22,88	58
38	M	186	95	27,46	50
39	N	158	42	16,824	
40	M	183	80	23,888	
41	M	179,5	78	24,208	52
42	M	183	85	25,381	55
43	N	164	58	21,565	
44	M	177	71	22,663	
45	M	177	75	23,939	56
46	N	165	46	16,896	
47	M	176	80	25,826	52
48	M	185	100	29,218	58
49	M	177	75	23,939	3
50	N	177	62	19,79	56
51	M	186	76	21,968	55
52	M	178	68	21,462	57
53	M	173	72	24,057	58
54	N	165	59	21,671	56
55	N	167	66	23,665	4
56	N	177	85	27,131	4
57	N	161	44	16,975	56
58	N	159	50	19,778	4
59	N	175	51	16,653	3
60	N	158	50	20,029	55
61	N	167	62,5	22,41	55
62	M	186	95	27,46	57
63	N	170	60	20,761	55
64	M	197	85	21,902	50
65	N	153	65	27,767	57
66	N	170	60	20,761	58
67	N	170	73	25,26	57,5

2. Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem? Kas kehamassiindeks on tugevamini seotud pikkusega või kehamassiga? Millise tunnusega on enam seotud peaümbermõõt?

Kirjeldage mõnda seost (pange vastavad laused ka kirja)!

Ülesande 3 tööjuhend

Illustreerige tunnuste 'Pikkus' ja 'KMI' vahelist seost hajuvusdiagrammiga (punktdiagrammiga).



- Pealkiri ja legend ✗
- Taust valgeks
- Ruudujooned ✗
- Pealkirjad x- ja y-teljele
- Telgede ühikud ja nimed kirjasuuruses 10
- Y-telje ulatus 15-33 märgiste vahega 5
- X-telje ulatus 150-200 märgiste vahega 10
- Sümbolid suuremaks (ja seest näiteks helesiniseks :)