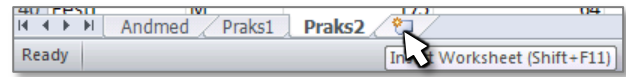


Biomeetria praks 3

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excel*'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht,
3. nimetage see ümber leheküljeks 'Praks3' ja
4. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks3' ülemisse vasakusse nurka.



Ülesanne 1.

- Leidke andmetabeli alla (NB! Vähemalt üks tühi rida jätke vahele!) kõigi arvutunnuste kohta vaatluste arv (n), keskmine väärtus (\bar{x}), mediaan, standardhälve (s), standardviga (se), minimaalne ja maksimaalne väärtus, kasutades *Exceli* funktsioone.
- Lisage andmetabelisse uus tunnus nimega 'KMI' (kehamassiindeks) ja arvutage selle väärtused kõigile tudengitele valemist

$$\text{KMI} = \text{Kehamass, kg} / (\text{Pikkus, m})^2.$$

Leidke eelnevalt nimetatud arvarakteristikute väärtused ka uuele tunnusele.

Tööjuhend

1. Jätke andmetabeli alla vähemalt üks tühi rida

(see on vajalik, et *Excel* mitmete operatsioonide teostamisel – näiteks andmete sorteerimisel või filtreerimisel või *Pivot Table*'i rakendamisel – ei tõlgendaks arvutatud keskmisi ja muid näitajaid andmetabeli osana)

ja kirjutage esimesse veergu leitavate arvarakteristikute nimed (siis on hiljem lihtsam aru saada, mida kuhugi arvutatud on).

	A	B	C
52	muu	M	189
53	Eesti	N	173
54			
55	Vaatluste arv		
56	Keskmine		
57	Mediaan		
58	Standardhälve		
59	Standardviga		
60	Min		
61	Max		

2. Arvutage kõigi arvarakteristikute väärtused tudengite pikkuse kohta, kasutades *Exceli* funktsioone.

a. Selleks võite valida vastava funktsiooni *Exceli* funktsioonide listist (vajalike funktsioonide nimed leiate järgmiselt leheküljelt punktist b):

	A	B	C	D
51	Eesti	N	168	62
52	muu	M	189	78
53	Eesti	N	173	68
54				
55	Vaatluste arv			
56	Keskmine			
57	Mediaan			
58	Standardhälve			
59	Standardviga			
60	Min			
61	Max			

Insert Function

Search for a function:

Type a brief description of what you want to do and then click Go

Or select a category: Statistical

Select a function:

- CORREL
- COUNT**
- COUNTA
- COUNTBLANK
- COUNTIF
- COVAR
- CRITBINOM

COUNT(value1;value2;...)
Counts the number of cells that contain numbers and numbers within the list of arguments.

[Help on this function](#)

OK Cancel

	A	B	C	D	E	F
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALA
2	Eesti	N	180	76	56	
3	Eesti	N	178	65	56	
4	Eesti	M	177	70	57	
5	Eesti	M	187	75	45	
6	Eesti	M	186	74	43	
7	Eesti	N	165	62	42	
8	Eesti	N	170	67	55,5	
9	Eesti	N	177	59	42	
10	Eesti	N	166	47	55	
11	Eesti	N	165	55	42	
12	Eesti	M	180	68	51	
13	Eesti	N	161	49	56	
14	Eesti	N	168	54	50	
15	Eesti	N	167	67	56	
16	Eesti	N	160	50	55	
17	Eesti	N	164	53	59	
18	Eesti	M	194	105	57	
19	Eesti	M	178	65	53	
20	Eesti	M	177	90	57	
21	Eesti	N	171	57	50	
22	Eesti	M	187	39	58	
23	Eesti	M	189	81	54	
24	Eesti	M	186	98	56	
25	Eesti	N	171	65	55	
26	Eesti	M	183	110	57	
27	Eesti	M	193	100	58	
28	Eesti	N	183	79	54	
29	Eesti	N	177	75	55	
30	Eesti	N	164	54	52	
31	Eesti	N	170	60	55	
32	Eesti	N	175	65	57	
33	Eesti	N	168	69	55	
34	Eesti	M	186	94	58	
35	Eesti	N	160	55	55	
36	Eesti	N	175	69	53	
37	Eesti	N	151	53	52	
38	Eesti	M	198	110	60	
39	Eesti	M	174	120	56	
40	Eesti	N	180	80	58	
41	Eesti	N	171	72		
42	Eesti	N	169	80	57	
43	Eesti	N	160	64	54	
44	Eesti	N	170	65	55	
45	Eesti	N	176	58	54	
46	Eesti	N	167	55	54	
47	Eesti	N	160	59	56	
48	Eesti	N	172	80	54	
49	Eesti	N	170	60		
50	Eesti	N	176	57	56	
51	Eesti	N	168	62	56	3
52	muu	M	189	78	57	
53	Eesti	N	173	68	45	
54						
55	Vaatluste arv		=COUNT(C2:C53)			
56	Keskmine					
57	Mediaan					

Function Arguments

COUNT

Value1: C2:C53 = {180;178;177;187;186;165;170;177...}

Value2: = number

= 52

Counts the number of cells in a range that contain numbers.

Value1: value1;value2;... are 1 to 255 arguments that can contain or refer to a variety of different types of data, but only numbers are counted.

Formula result = 52

[Help on this function](#)

OK Cancel

Vaatluste arv	52
Keskmine	
Mediaan	
Standardhälve	
Standardviga	
Min	
Max	

- b. Teades funktsiooni nime ja süntaksit, võite trükkida vastava valemi ka kohe *Exceli* töölehe vastavasse lahtrisse.
(NB! Ärge unustage alustamast valemit võrdusmärgiga '='!)

Kõik need funktsioonid on rakendatavad ka eelmisel leheküljel esitatud viisil – valige ise, milline variant omale arusaadavam ja mugavam tundub (proovige nii üht kui ka teist varianti).

- c. Et *Excelis* puudub eraldi funktsioon **standardvea** leidmiseks, tuleb arvutused teostada, tuginedes standardvea arvutusvalemile

$$se = s/\sqrt{n}$$

(st, et vastav valem tuleb ise sisestada):

Vaatluste arv	=COUNT(C2:C53)
Keskmine	=AVERAGE(C2:C53)
Mediaan	=MEDIAN(C2:C53)
Standardhälve	=STDEV.S(C2:C53)
Standardviga	
Min	=MIN(C2:C53)
Max	=MAX(C2:C53)

	A	B	C	D
54				
55	Vaatluste arv		52	
56	Keskmine		174,3	
57	Mediaan		173,5	
58	Standardhälve		10,0	
59	Standardviga		=C58/SQRT(C55)	
60	Min		151	
61	Max		198	

3. Rakendage samu funktsioone ning arvutage soovitud arvarakteristikute väärtused kõigi andmestikus sisalduvate arvtunnuste jaoks.

52	muu	M	189	78	57	47	5	3
53	Eesti	N	173	68	45	40	1	4
54								
55	Vaatluste arv		52					
56	Keskmine		174,3					
57	Mediaan		173,5					
58	Standardhälve		10,0					
59	Standardviga		1,4					
60	Min		151					
61	Max		198					

4. Kui *Excel* seda automaatselt ei teinud, siis **ümardage keskmised, standardhälbed ja standardvead ühe kohani peale koma.**

Tulemus:

	A	B	C	D	E	F	G	H
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEND	MAT_HINNE
54								
55	Vaatluste arv		52	52	50	52	52	52
56	Keskmine		174,3	71,2	53,9	40,8	1,9	3,7
57	Mediaan		173,5	67	55	40	1,5	4
58	Standardhälve		10,0	17,4	4,5	3,2	1,5	0,7
59	Standardviga		1,4	2,4	0,6	0,4	0,2	0,1
60	Min		151	47	42	36	0	3
61	Max		198	120	60	47	8	5

5. Kirjutage mõned laused uuritud tunnuste väärtuste paiknemise ja varieeruvuse kohta.

6. Lisage andmetabelisse kehamassi veeru järele tühi veerg, kirjutage esimesse lahtrisse tunnuse nimeks 'KMI' (kehamassiindeks) ja arvutage kehamassiindeksi väärtused kõigile tudengitele valemist

$$\text{KMI} = \text{Kehamass, kg} / (\text{Pikkus, m})^2.$$

	A	B	C	D	E	F
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P
2	Eesti	N	180	76	=D2/((C2/100)^2)	
3	Eesti	N	178	65		

	A	B	C	D	E	F
Eesti	N		180	76	23,457	
Eesti	N		178	65		
Eesti	M		177	70		
Eesti	M		187	75		
Eesti	M		186	74		
Eesti	N		165	62		
Eesti	N		170	67		
Eesti	N		177	59		
Eesti	N		166	47		
Eesti	N		165	55		
Eesti	M		180	68		
Eesti	N		161	49		
Eesti	N		168	54		
Eesti	N		167	67		
Eesti	N		160	50		
Eesti	N		164	53		
Eesti	M		194	105		
Eesti	M		178	65		
Eesti	M		177	90		
Eesti	N		171	57		
Eesti	M		187	99		
Eesti	M		189	81		
Eesti	M		186	98		
Eesti	N		171	65		
Eesti	M		183	110		
Eesti	M		193	100		
Eesti	N		183	79		
Eesti	N		177	75		
Eesti	N		164	54		
Eesti	N		170	60		
Eesti	N		175	65		
Eesti	N		168	69		
Eesti	M		186	94		
Eesti	N		160	55		
Eesti	N		175	69		
Eesti	N		151	53		
Eesti	M		198	110		
Eesti	M		174	120		
Eesti	N		180	80		
Eesti	N		171	72		
Eesti	N		169	80		
Eesti	N		160	64		
Eesti	N		170	65		
Eesti	N		176	58		
Eesti	N		167	55		
Eesti	N		160	59		
Eesti	N		172	80		
Eesti	N		170	60		
Eesti	N		176	57		
Eesti	N		168	62		
Eesti	M		189	78		
Eesti	N		173	68		

NB! Jälgi sulgude arvu ja paigutust!

Kas te saate sellest *Exceli* valemist aru?

- Astendamise märk ^ on enamasti saadav klahvikombinatsiooni 'AltGr' + 'Ä' tulemusel.
- Alternatiiv mingi suuruse ruutu võtmiseks on korrutada see iseendaga: (C2/100)*(C2/100);
- kasutada võib ka *Exceli* üldist astendamise funktsiooni POWER(C2/100;2) – siin esimene argument on astendatav ja teine astendaja.

7. Leidke vajalikud arvarakteristikute väärtused ka uuele tunnusele.

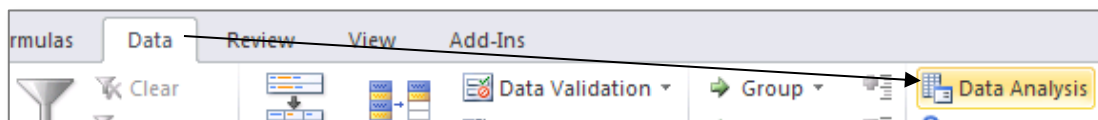
Eesti	N	173	63	22,72044	45
Vaatluste arv		52	52	52	50
Keskmine		174,3	71,2	23,2	53,9
Mediaan		173,5	67	22,51098	55
Standardhälve		10,0	17,4	4,0	4,5
Standardviga		1,4	2,4	0,6	0,6
Min		151	47	17,05618	42
Max		198	120	39,63535	60

Ülesanne 2.

- Arvutage tudengite pikkuse, massi, kehamassiindeksi, peaümbermõõdu ja jalanumbri kohta nii palju arvarakteristikuid, kui protseduur *Descriptive Statistics* (*Data*-sakk → *Data Analysis*...) võimaldab.
- Leidke ka 90%, 95% või 99% usalduspiirid keskmistele väärtustele. Mida need usalduspiirid näitavad?

Tööjuhend

- Arvarakteristikute arvutamine: *Data*-sakk → *Data Analysis*... → *Descriptive Statistics*



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P	JALANR	ODE_VEI	MAT_HIN	HOMMIK	PUDER	LEMMIK	HAIGE	SPORT	SUUSK	AUTO	OLU	SUITS	TEATER	KIND		
2	Eesti	N	180	76	23,457	56	42	2	3	muu	jah	ei	ei	jah	jah	jah	0	ei	viimase a	viimase kuu	jooksul	
3	Eesti	N	178	65	20,515	56	39	4									0,25	ei	rohkem k	viimase aasta	jooksul	
4	Eesti	M	177	70	22,344	57	42	3									5	jah	viimase a	viimase aasta	jooksul	
5	Eesti	M	187	75	21,448	45	45	3									2	jah	viimase a	viimase aasta	jooksul	
6	Eesti	M	186	74	21,39	43	44	1									2	jah	viimase a	viimase kuu	jooksul	
7	Eesti	N	165	62	22,773	42	37	3									0	ei	viimase a	viimase kuu	jooksul	
8	Eesti	N	170	67	23,183	55,5	40	4									0	ei	viimase a	viimase aasta	jooksul	
9	Eesti	N	177	59	18,832	42	39	1									0	ei	viimase a	viimase aasta	jooksul	
10	Eesti	N	166	47	17,056	55	38	2														
11	Eesti	N	165	55	20,202	42	38	1														
12	Eesti	M	180	68	20,988	51	44	0														
13	Eesti	N	161	49	18,904	56	37	2														
14	Eesti	N	168	54	19,133	50	40	0														
15	Eesti	N	167	67	24,024	56	40	2														
16	Eesti																					
17	Eesti																					
18	Eesti																					
19	Eesti																					
20	Eesti																					
21	Eesti																					
22	Eesti	M	187	99	28,311	58	45	1														
23	Eesti	M	189	81	22,676	54	46	1														
24	Eesti	M	186	98	28,327	56	45	0														
25	Eesti	N	171	65	22,229	55	38	1														
26	Eesti	M	183	110	32,847	57	43	0														
27	Eesti	M	193	100	26,846	58	46,5	1														
28	Eesti	N	183	79	23,9																	
29	Eesti	N	177	75	23,9																	
30	Eesti	N	164	54	20,0																	
31	Eesti	N	170	60	20,1																	
32	Eesti	N	175	65	21,2																	
33	Eesti	N	168	69	24,4																	
34	Eesti	M	186	94	27,171	58	46	1														
35	Eesti	N	160	55	21,484	55	36	2														
36	Eesti	N	175	69	22,531	53	39	8														
37	Eesti	N	151	53	23,245	52	36	3														
38	Eesti	M	198	110	28,058	60	47	4														
39	Eesti	M	174	120	39,635	56	45	3														
40	Eesti	N	180	80	24,691	58	42	2														
41	Eesti	N	171	72	24,623		41	3														
42	Eesti	N	169	80	28,01	57	40	1														
43	Eesti	N	160	64	25	54	38	0														
44	Eesti	N	170	65	22,491	55	39	1														
45	Eesti	N	176	58	18,724	54	40	1														
46	Eesti	N	167	55	19,721	54	39	4														
47	Eesti	N	160	59	23,047	56	38	1														
48	Eesti	N	172	80	27,042	54	39	1														
49	Eesti	N	170	60	20,761		39	2														
50	Eesti	N	176	57	18,401	56	41	1														
51	Eesti	N	168	62	21,967	56	38,5	2														
52	muu	M	189	78	21,836	57	47	5														
53	Eesti	N	173	68	22,72	45	40	1														

Võimalik on analüüsida mitut tunnust korraga tingimusel, et nende väärtused paiknevad kõrvuti veergudes

Valik 'Labels in first row' peab olema märgitud, kui andmed on ette antud koos esimeses reas paikneva nimega.

Lisavalikute 'Summary statistics' jt kohta vt järgmine lk.

Väljundtabeli vasaku ülemise nurga asukoht

Selgitus protseduuri *Descriptive Statistics* lisavalikutest eelmisel lehel:

- valiku *Summary statistics* tulemusena arvutab *Excel* kaheistkümmne põhilise arvarakteristiku väärtused;
- valiku *Confidence Level for Mean: 95%* tulemusena arvutatakse suurus, mis tuleb keskmisele juurde liita või lahutada, saamaks ülemist ja alumist usalduspiiri; vaikumisi kasutatava 95% asemele võib ise trükkida mõne teise arvu (näiteks 90 või 99);
- valikute *Kth Largest* ja *Kth Smallest* tulemusena väljastatakse järjekorranumbriga *K* väärtus vastavalt suurima ja vähima väärtuse poolt lugedes;
Kth Largest = 1 korral on tulemuseks maksimaalne väärtus ja *Kth Smallest* = 1 korral minimaalne väärtus.
 Et aga miinimum ja maksimum sisalduvad ka valiku *Summary statistics* väljundis, on antud juhul mõistlik tellida näiteks suuruselt järgmised väärtused (siis *K* = 2).

• Tulemus:

PIKKUS	MASS	KMI	PEA_P	JALANR					
Mean	174,269	Mean	71,1923	Mean	23,203	Mean	53,87	Mean	40,7885
Standard I	1,39167	Standard I	2,40976	Standard I	0,55793	Standard I	0,63869	Standard Error	0,44032
Median	173,5	Median	67	Median	22,511	Median	55	Median	40
Mode	177	Mode	65	Mode	20,5151	Mode	56	Mode	39
Standard I	10,0355	Standard I	17,3771	Standard I	4,02328	Standard I	4,5162	Standard Deviation	3,17519
Sample Va	100,71	Sample Va	301,962	Sample Va	16,1868	Sample Va	20,396	Sample Variance	10,0818
Kurtosis	-0,23785	Kurtosis	0,52817	Kurtosis	4,43715	Kurtosis	1,83024	Kurtosis	-0,80794
Skewness	0,24427	Skewness	1,05584	Skewness	1,64878	Skewness	-1,58938	Skewness	0,61252
Range	47	Range	73	Range	22,5792	Range	18	Range	11
Minimum	151	Minimum	47	Minimum	17,0562	Minimum	42	Minimum	36
Maximum	198	Maximum	120	Maximum	39,6354	Maximum	60	Maximum	47
Sum	9062	Sum	3702	Sum	1206,55	Sum	2693,5	Sum	2121
Count	52	Count	52	Count	52	Count	50	Count	52
Largest(2)	194	Largest(2)	110	Largest(2)	32,8466	Largest(2)	59	Largest(2)	47
Smallest(2)	160	Smallest(2)	49	Smallest(2)	18,4013	Smallest(2)	42	Smallest(2)	36
Confidenc	2,79389	Confidenc	4,83781	Confidenc	1,12009	Confidenc	1,28349	Confidence Level(95,0%)	0,88398

Valiku *Summary statistics* tulemus

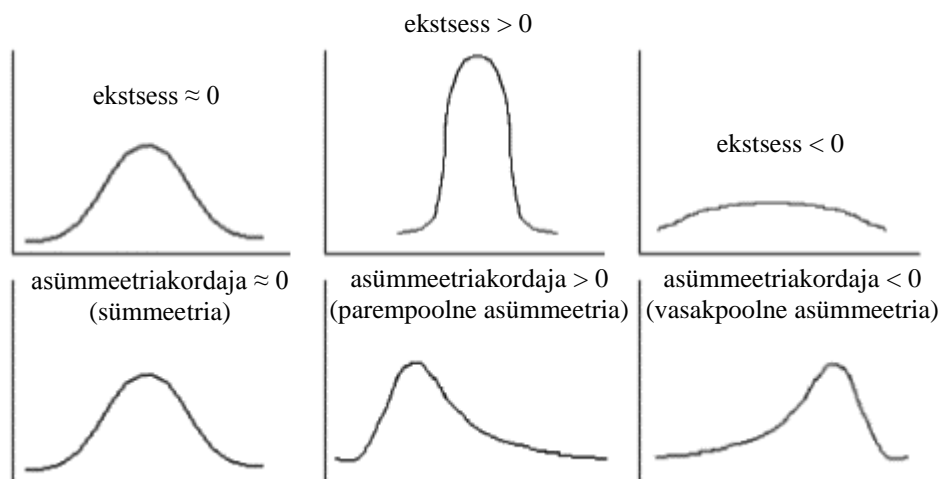
• **Lisalugemine – uuritava tunnuse jaotuse kuju iseloomustamine**

Enamusest protseduuri *Descriptive Statistics* väljundis sisalduvatest arvarakteristikutest on ennegi juttu olnud.

Siiski on siin ka kaks uut suurust, mida kasutatakse peamiselt uuritava tunnuse jaotuse kuju iseloomustamiseks – need suurused on **ekstsess** ehk **järsakuskordaja** (ingl *kurtosis*) ja **asümmeetriakordaja** (ingl *skewness*). Sellest, mida need karakteristikud mõeldavad, annavad parema ettekujutuse järgnevad joonised:

Jaotuse märkimisväärselt erinevusest normaaljaotusest on mõtet rääkida siis, kui ükskõik kumb neist kordajatest omandab absoluutväärtuselt 1-st suurema väärtuse ...

Eriti palju neid kordajaid siiski ei kasutata.



Antud juhul on näiteks kehamassiindeksile vastav järsakuskordaja (4,44) märksa suurem võrreldes teiste kehamõõtude sama parameetriga – see näitab, et enamuse tudengite kehamassiindeksi väärtused jäävad küllalt kitsasse vahemikku, samas leiduvad üksikud teistest palju suurema või väiksema kehamassiindeksi väärtusega tudengid.

- **Jaotuse sümmeetrilisuse üle otsustamisel kasutatakse sageli (asümmeetriakordaja asemel) keskmise ja mediaani võrdlust.**

Nimelt, kuna aritmeetiline keskmine on tundlik erandlike väärtuste suhtes, siis vihjab

$\bar{x} > med$ sellele, et jaotuse kuju on parempoolse asümmeetriaga (leiduvad üksikud teistest palju suuremad väärtused, ja seega asümmeetriakordaja > 0),

$\bar{x} < med$ aga sellele, et jaotuse kuju on vasakpoolse asümmeetriaga (leiduvad üksikud teistest palju väiksemad väärtused, ja seega asümmeetriakordaja < 0).

- **Vaadake, kas kirjeldatud seos keskmise ja mediaani erinevuse ning asümmeetriakordaja väärtuse vahel peab paika ka teie kursuse tudengite kehamõõtude puhul.**

2. Leidke 90%, 95% või 99% usalduspiirid keskmisele väärtustele. Mida need näitavad?

Kuna *Excel* ise usalduspiire välja ei arvuta, tuleb need enesest leida.

Selleks võib protseduuri *Descriptive Statistics* väljundtabelit täiendada kahe reaga, kuhu tuleks selguse huvides ka kirja panna, mida uued arvutatavad suurused enesest kujutavad.

	V	W
1	PIKKUS	
2		
3	Mean	174,269
4	Standard Error	1,39167
5	Median	173,5
6	Mode	177
7	Standard Deviation	10,0355
8	Sample Variance	100,71
9	Kurtosis	-0,23785
10	Skewness	0,24427
11	Range	47
12	Minimum	151
13	Maximum	198
14	Sum	9062
15	Count	52
16	Largest(2)	194
17	Smallest(2)	160
18	Confidence Level(95,0%)	2,79389
19		
20	Alumine 95% usalduspiir	=W3-W18
21	Ülemine 95% usalduspiir	=W3+W18

Usalduspiirid keskmisele leitakse valemist

$$\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

Excel väljastab toodud valemi mõlemad liidetavad, mistap on edasine usalduspiiride välja arvutamine lihtne.

Alumine 95% usalduspiir	171,475
Ülemine 95% usalduspiir	177,063

Seega, tõlgendades antud andmestikku kui valimit Maaülikooli esmakursuslastest õppeaastal 2012-2013, võib väita, et esimese kursuse tudengite **keskmine pikkus** jääb 95%-lise tõenäosusega vahemikku 171,5 cm kuni 177,1 cm. St, et mõõtes ära **kõigi** Maaülikooli esimese kursuse tudengite pikkused ja arvutades keskmise, peaks saadud tegelik keskmine 95%-lise tõenäosusega jääma leitud piiridesse.

- Kui keegi leidis 95% usalduspiiride asemel 90% või 99% usalduspiirid, siis need peaksid tulema vastavalt (171,9; 176,6) ja (170,5; 178,0). **Miks on 90%-line usaldusintervall kitsam?**
- **Arvutage usaldusintervall ka teiste tunnuste keskmisele väärtustele ning püüdke neist vähemalt ühe kohta sõnastada lõppjärelus!!**