

Biomeetria praks 3

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

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

Ülesanne 1.

- Leidke andmetabeli alla (NB! Vähemalt üks tühi rida jätke vahele!) kõigi arvtunnuste 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 $KMI = \text{Kehamass, kg} / (\text{Pikkus, m})^2$. 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, *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).

47	Eesti	MI	105	62	59	45
48	Eesti	N	174	60	49	39
49						
50	Vaatluste arv					
51	Keskmine					
52	Mediaan					
53	Standardhälve					
54	Standardviga					
55	Min					
56	Max					

2. Arvutage kõigi arvarakteristike 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	E
46	Eesti	N		70	55
47	Eesti	M	185	82	59
48	Eesti	N	174	60	49
49					
50	Vaatluste arv				
51	Keskmine				
52	Mediaan				
53	Standardhälve				
54	Standardviga				
55	Min				
56	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	G
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEI MA
2	Eesti	N	170	64	57	38	1
3	Eesti	N	170	60	57	40	0
4	Eesti	N	163	64	54	38	2
5	Eesti	N	161,5	52			
6	Eesti	N	171	75			
7	Soome	N	178	70			
8	Eesti	N	175	67			
9	Eesti	N	160	52			
10	Eesti	N	160	55			
11	Eesti	N	170	69			
12	Eesti	N	167	57			
13	Eesti	N	172	59			
14	Eesti	N	167	60			
15	Eesti	N	175	63			
16	Eesti	N	173	73			
17	Eesti	N	168	65			
18	Soome	N	170	80			
19	Eesti	N	165	70			
20	Soome	N	170	60			
21	Eesti	N	175	60			
22	Eesti	N	160	50			
23	Eesti	N	160	51			
24	Eesti	N	167	55			
25	Eesti	N	161	54			
26	Eesti	N	165	54,3	54	38	2
27	Eesti	N	158	55	55	35	0
28	Soome	N	173	62	54	39	1
29	Eesti	N	174	58	54	39	2
30	Eesti	M	175	76	57	43	2
31	Eesti	N	180	65	55	40	1
32	Eesti	N	176	64	57	40	0
33	Soome	N	164	63	42	38	2
34	Soome	N	172	70	43,5	39	2
35	Eesti	N	161	56	55	38	1
36	Eesti	N	181	60	55	40	1
37	Eesti	N	164	55	53	38	1
38	Eesti	N	177	71	57	41	2
39	Eesti	N	175	70	54	41	1
40	Eesti	N	170	110	57	43	0
41	Eesti	N	164	58	54	38,5	1
42	Eesti	N	163	50,5	55,5	37	3
43	Eesti	N	159	57	56	37	4
44			163	54	52	38	1
45	Eesti	N	175	55	55	38,5	1
46	Eesti	N	160	70	55	39	2
47	Eesti	M	185	82	59	45	1
48	Eesti	N	174	60	49	39	2
49							
50	Vaatluste arv		=COUNT(C2:C48)				
51	Keskmine						
52	Mediaan						
53	Standardhälve						
54	Standardviga						
55	Min						
56	Max						

Function Arguments

COUNT

Value1: C2:C48 = {170;170;163;161,5}

Value2: = number

= 47

Counts the number of cells that contain numbers and numbers within the list of arguments.

Value1: value1;value2;... are 1 to 30 arguments that can contain or refer to a variety of different types of data, but only numbers are counted.

Formula result = 47

[Help on this function](#)

OK Cancel

50	Vaatluste arv	47
51	Keskmine	
52	Mediaan	
53	Standardhälve	
54	Standardviga	
55	Min	
56	Max	

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

Kõik need funktsioonid on rakendatavad ka eelmisel leheküljel esitatud viisil – valige ise, milline variant omale arusaadavam ja mugavam tundub.

Vaatluste arv	=COUNT(C2:C48)
Keskmine	=AVERAGE(C2:C48)
Mediaan	=MEDIAN(C2:C48)
Standardhälve	=STDEV(C2:C48)
Standardviga	
Min	=MIN(C2:C48)
Max	=MAX(C2:C48)

- 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):

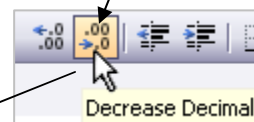
	A	B	C	D
49				
50	Vaatluste arv		47	
51	Keskmine		168,9	
52	Mediaan		170	
53	Standardhälve		6,7	
54	Standardviga		=C53/SQRT(C50)	
55	Min		158	
56	Max		185	

3. Rakendage samu funktsioone, arvutamaks soovitud arvkarakteristikute väärtused kõigi andmestikus sisalduvate arvtunnuste jaoks.

47	Eesti	M	185	82	59	45	1	4	m
48	Eesti	N	174	60	49	39	2	4	pt
49									
50	Vaatluste arv		47						
51	Keskmine		168,861702						
52	Mediaan		170						
53	Standardhälve		6,721722						
54	Standardviga		0,980464						
55	Min		158						
56	Max		185						

Ümardage keskmised, standardhälbed ja standardvead ühe kohani peale koma.

Vaatluste arv	47	47	47	47	47	47
Keskmine	168,861702	62,995745	54,404255	38,829787	1,489362	4,085106
Mediaan	170	60	55	38,5	1	4
Standardhälve	6,721722	10,586024	3,412952	1,868726	0,952624	0,619600
Standardviga	0,980464	1,544130	0,497830	0,272582	0,138955	0,090378
Min	158	50	42	35	0	3
Max	185	110	59	45	4	5



Vaatluste arv	47	47	47	47	47	47
Keskmine	168,9	63,0	54,4	38,8	1,5	4,1
Mediaan	170	60	55	38,5	1	4
Standardhälve	6,7	10,6	3,4	1,9	1,0	0,6
Standardviga	1,0	1,5	0,5	0,3	0,1	0,1
Min	158	50	42	35	0	3
Max	185	110	59	45	4	5

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

5. Lisage andmetabelisse kehamassi veeru järele tühi veerg, kirjutage esimesse lahtrisse tunnuse nimeks 'KMI' (kehamassiindeks) ja arvutage kehamassiindeksi väärtused kõigile tudengitele valemist $KMI = \text{Kehamass, kg} / (\text{Pikkus, m})^2$.

	A	B	C	D	E	F
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P JAI
2	Eesti	N	170	64	=D2/((C2/100)^2)	
3	Eesti	N	170	60		57

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.

	A	B	C	D	E	F
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P
2	Eesti	N	170	64	22,1453	57
3	Eesti	N	170	60		57
4	Eesti	N	163	64		57
5	Eesti	N	161,5	52		57
6	Eesti	N	171	75		55,3
7	Soome	N	178	70		57
8	Eesti	N	175	67		57
9	Eesti	N	160	52		50
10	Eesti	N	160	55		50
11	Eesti	N	170	69		57
12	Eesti	N	167	57		57
13	Eesti	N	172	59		55,3
14	Eesti	N	167	60		49
15	Eesti	N	175	63		57
16	Eesti	N	173	73		57
17	Eesti	N	168	65		57
18	Soome	N	170	80		57,3
19	Eesti	N	165	70		50
20	Soome	N	170	60		55,3
21	Eesti	N	175	60		57
22	Eesti	N	160	50		50
23	Eesti	N	160	51		50
24	Eesti	N	167	55		50
25	Eesti	N	161	54		57
26	Eesti	N	165	54,3		57
27	Eesti	N	168	55		57
28	Soome	N	173	62		57
29	Eesti	N	174	58		57
30	Eesti	M	175	76		57
31	Eesti	N	180	65		57
32	Eesti	N	176	64		57
33	Soome	N	164	63		49
34	Soome	N	172	70		43,3
35	Eesti	N	161	56		57
36	Eesti	N	181	60		57
37	Eesti	N	164	55		57
38	Eesti	N	177	71		57
39	Eesti	N	175	70		57
40	Eesti	N	170	110		57
41	Eesti	N	164	58		57
42	Eesti	N	163	50,5		55,3
43	Eesti	N	159	57		57
44			163	54		57
45	Eesti	N	175	55		57
46	Eesti	N	160	70		57
47	Eesti	M	185	82		57
48	Eesti	N	174	60		49

6. Leidke vajalikud arvkarakteristikute väärtused ka uuele tunnusele.

	A	B	C	D	E	F	G	H
48	Eesti	N	Kopeeri / Copy -> Kleebi / Paste		39	2	4	
49								
50	Vaatluste arv	47	47	47	47	47	47	47
51	Keskmine	168,9	63,0	22,1	54,4	38,8	1,5	4,1
52	Mediaan	170	60	21,48438	55	38,5	1	4
53	Standardhälve	6,7	10,6	3,3	3,4	1,9	1,0	0,6
54	Standardviga	1,0	1,5	0,5	0,5	0,3	0,1	0,1
55	Min	158	50	17,95918	42	35	0	3
56	Max	185	110	38,06228	59	45	4	5

Ülesanne 2.

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

Tööjuhend

- Arvarakteristikute arvutamine: *Tools / Tööriistad -> Data Analysis ... -> Descriptive Statistics*

	A	B	C	D	E	F	G	H	I	J	S	T	U	V	W
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P	JALANR	ODE	VEN	MAT_HIN	HOMMIK	TEATER	KIND		
2	Eesti	N	170	64	22,1453	57	38	1	4	helbed või viimase 10 päeva jooksul					
3	Eesti	N	170	60	20,7612	57	40	0	4	võileib viimase 10 päeva jooksul					
4	Eesti	N	163	64	24,0882	54				uu jooksul					
5	Eesti	N	161,5	52	19,9369	54				uu jooksul					
6	Eesti	N	171	75	25,6489	55,5				uu jooksul					
7	Soome	N	178	70	22,0932	52				10 päeva jooksul					
8	Eesti	N	175	67	21,8776	57									
9	Eesti	N	160	52	20,3125	56	38								
10	Eesti	N	160	55	21,4844	56	36								
11	Eesti	N	170	69	23,8754	52	36								
12	Eesti	N	167	57	20,4382	57	38,5								
13	Eesti	N	172	59	19,9432	55,5	38								
14	Eesti	N	167	60	21,5139	45	39								
15	Eesti	N	175	63	20,5714	57	40								
16	Eesti	N	173	73	24,3911	52	41								
17	Eesti	N													
18	Soome	N													
19	Eesti	N													
20	Soome	N													
21	Eesti	N													
22	Eesti	N	160	50	19,5313	56	37								
23	Eesti	N	160	51	19,9219	56	36								
24	Eesti	N	167	55	19,721	56	37								
25	Eesti	N	161	54	20,8325	54	38								
26	Eesti	N	165	54,3											
27	Eesti	N	158	55											
28	Soome	N	173	62											
29	Eesti	N	174	58											
30	Eesti	M	175	76											
31	Eesti	N	180	65	20,0617	55	40								
32	Eesti	N	176	64	20,6612	57	40								
33	Soome	N	164	63	23,4236	42	38								
34	Soome	N	172	70	23,8614	43,5	39								
35	Eesti	N	161	56	21,6041	55	38	1	4	puder	rohkem kui viimase aasta jooksul				
36	Eesti	N	181	60	18,3145	55	40	1	4	võileib	viimase 10 päeva jooksul				
37	Eesti	N	164	55	20,4491	53	38	1	4	puder	viimase 10 päeva jooksul				
38	Eesti	N	177	71	22,6627	57	41	2	4	võileib	viimase 10 päeva jooksul				
39	Eesti	N	175	70	22,8571	54	41	1	5	ei söö tavaliselt	viimase 10 päeva jooksul				
40	Eesti	N	170	110	38,0623	57	43	0	4	muu	viimase 10 päeva jooksul				
41	Eesti	N	164	58	21,5645	54	38,5	1	3	võileib	viimase 10 päeva jooksul				
42	Eesti	N	163	50,5	19,0071	55,5	37	3	4	võileib	rohkem kui rohkem kui aasta tagasi				
43	Eesti	N	159	57	22,5466	56	37	4	4	võileib	viimase 10 päeva jooksul				
44			163	54	20,3244	52	38	1	5	helbed või	rohkem kui viimase aasta jooksul				
46	Eesti	N	175	55	17,9592	55	38,5	1	5	ei söö tavaliselt	viimase 10 päeva jooksul				
46	Eesti	N	160	70	27,3438	55	39	2	4	võileib	rohkem kui viimase aasta jooksul				
47	Eesti	M	185	82	23,9591	59	45	1	4	muu	viimase 10 päeva jooksul				
48	Eesti	N	174	60	19,8177	49	39	2	4	puder	rohkem kui viimase aasta jooksul				

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.

Descriptive Statistics

Input Range: \$C\$1:\$G\$48

Grouped By: Columns

Labels in first row:

Output options

Output Range: \$V\$1

Summary statistics:

Confidence Level for Mean: 95 %

Kth Largest: 2

Kth Smallest: 2

Väljundtabeli vasaku ülemise nurga asukoht

Selgitus protseduuri *Descriptive Statistics* lisavalikutest eelmisel lehel:

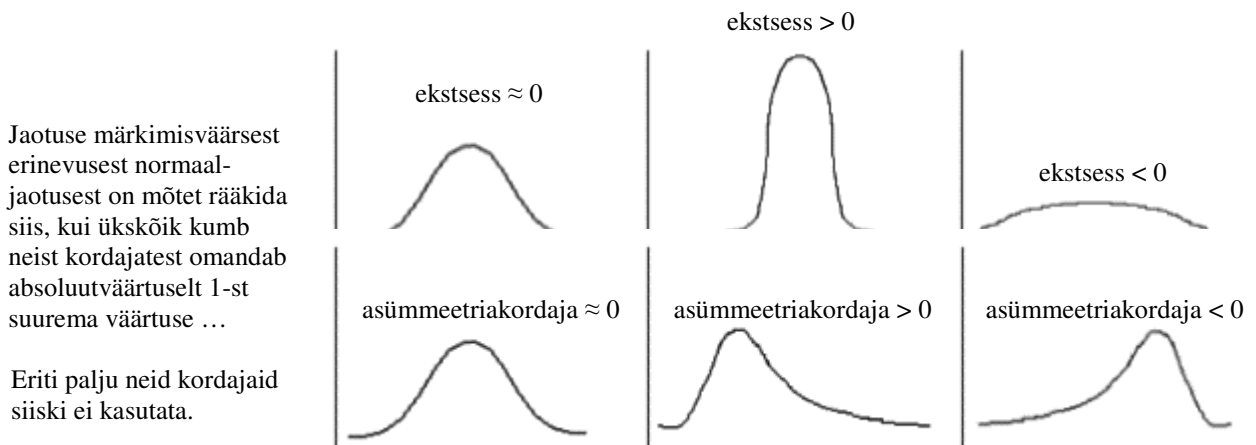
- valiku *Summary statistics* tulemusena arvutab *Excel* kaheteistkümne põhilise arvarakteristiku väärtused;
- valiku *Confidence Level for Mean: 95%* tulemusena arvutatakse suurus, mis tuleb keskmisele juurde liita või lahutada, saamaks ülemist ja alumist usalduspiiri; vaikimisi kasutatava 95% asemele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

• Tulemus:

PIKKUS	MASS	KMI	PEA_P	JALANIR							
Mean	168,862	Mean	62,9957	Mean	22,0598	Mean	54,4043	Mean	38,8298	} Valiku Summary statistics tulemus:	Keskmine
Standard	0,98046	Standard	1,54413	Standard	0,47575	Standard	0,49783	Standard Error	0,27258		Standardviga
Median	170	Median	60	Median	21,4844	Median	55	Median	38,5		Mediaan
Mode	170	Mode	60	Mode	20,7612	Mode	57	Mode	38		Mood
Standard	6,72172	Standard	10,586	Standard	3,26156	Standard	3,41295	Standard Deviation	1,86873		Standardhälve
Sample \	45,1815	Sample \	112,064	Sample \	10,6378	Sample \	11,6482	Sample Variance	3,49214		Dispersioon
Kurtosis	-0,7781	Kurtosis	7,29014	Kurtosis	11,8441	Kurtosis	5,48905	Kurtosis	2,06885		Ekstsess e järsakuskordaja
Skewness	0,20526	Skewness	2,07937	Skewness	2,79504	Skewness	-2,2447	Skewness	0,92379		Asümmeetriakordaja
Range	27	Range	60	Range	20,1031	Range	17	Range	10		Ulatus = Max - Min
Minimum	158	Minimum	50	Minimum	17,9592	Minimum	42	Minimum	35		
Maximum	185	Maximum	110	Maximum	38,0623	Maximum	59	Maximum	45		
Sum	7936,5	Sum	2960,8	Sum	1036,81	Sum	2557	Sum	1825		
Count	47	Count	47	Count	47	Count	47	Count	47	Vaatluste arv	
Largest(2)	181	Largest(2)	82	Largest(2)	27,6817	Largest(2)	57,5	Largest(2)	43		
Smallest(2)	159	Smallest(2)	50,5	Smallest(2)	18,3145	Smallest(2)	43,5	Smallest(2)	36		
Confiden	1,97357	Confiden	3,10817	Confiden	0,95763	Confiden	1,00208	Confidence Level(95,0%)	0,54868		

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

Enamustest protseduuri *Descriptive Statistics* väljundis sisalduvatel arvarakteristikutele on ennegi juttu olnud. Siiski on siin ka kaks uut suurust, mida kasutatakse peamiselt uuritava tunnuse jaotuse kuju iseloomustamiseks – need suurused on **ekstsess e järsakuskordaja** (ingl *kurtosis*) ja **asümmeetriakordaja** (ingl *skewness*). Sellest, mida need karakteristikud mõeldavad, annavad parema ettekujutuse järgnevad joonised:



- **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 ebasümmeetriaga (leiduvad üksikud teistest palju suuremad väärtused, ja seega asümmeetriakordaja > 0),

$\bar{x} < med$ aga sellele, et jaotuse kuju on vasakpoolse ebasümmeetriaga (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 keskmistele väärtustele. Mida need näitavad?

Kuna *Excel* ise usalduspiire välja ei arvuta, tuleb need eneselt 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	168,8617
4	Standard Error	0,980464
5	Median	170
6	Mode	170
7	Standard Deviation	6,721722
8	Sample Variance	45,18154
9	Kurtosis	-0,77811
10	Skewness	0,205265
11	Range	27
12	Minimum	158
13	Maximum	185
14	Sum	7936,5
15	Count	47
16	Largest(2)	181
17	Smallest(2)	159
18	Confidence Level(95,0%)	1,973571
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, mille alusel on lihtne mõlemad usalduspiirid välja arvutada.

Alumine 95% usalduspiir	166,8881
Ülemine 95% usalduspiir	170,8353

Seega, tõlgendades antud andmestikku kui valimit kogu teie kursusest, võib väita, et veterinaarmeditsiini eriala esimese kursuse tudengite keskmine pikkus jääb 95% tõenäosusega vahemikku 166,9 cm kuni 170,8 cm. St, et mõõtes ära **kõigi** esmakursuslaste pikkused ja arvutades keskmise, peaks saadud tegelik keskmine 95% tõenäosusega jääma leitud piiridesse.

- Kui keegi leidis 95% usalduspiiride asemel 90% või 99% usalduspiirid, siis need peaksid tulema vastavalt (167,2; 170,5) ja (166,2; 171,5). **Miks on 90% usaldusintervall kitsam?**
- **Arvutage usaldusintervall ka teiste tunnuste keskmistele väärtustele ning püüdke neist vähemalt ühe kohta sõnastada lõppjärelus!!**