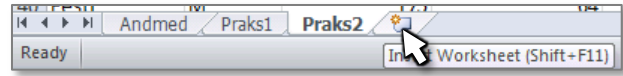


## Biomeetria praks 3

### 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 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).

57	Eesti	N	160	49	52	37
58						
59						
60	Vaatluste arv					
61	Keskmine					
62	Mediaan					
63	Standardhälve					
64	Standardviga					
65	Min					
66	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	E
56	Eesti	N	168	57
57	Eesti	N	168	60
58	Eesti	N	160	49
59				
60	Vaatluste arv			
61	Keskmine			
62	Mediaan			
63	Standardhälve			
64	Standardviga			
65	Min			
66	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	H	I	J	K	
1	RIIK	SUGU	PIKKUS	MASS	FEA_P	JALANR	ODE_VENI	MAT_HINI	HOMMIK	PUDER	LEMMIK	HA
2	Eesti	N	170	63	57	39	1	5	võlleib	jah	ei	ei
3	Eesti	N	161	57	59	37,5	2	4	võlleib	jah	jah	jah
4	Eesti	M	183	80	48	43	2	3	puder	jah	ei	ei
5	Eesti	N	183	81	58	42	2	5	võlleib	jah	jah	jah
6	Eesti	M	174	74	56	43	0	4	võlleib	nii ja nee	jah	ei
7	Eesti	N	171	70	57	39	3	4	puder	jah	ei	ei
8	Eesti	N	161	59	55	37	1	5	võlleib	nii ja nee	jah	jah
9	Eesti	N	172	80	56	44	2	4	võlleib	nii ja nee	jah	ei
10	Eesti	N	183,5	68	53,5	40	3	4	ei söö teav	nii ja nee	jah	jah
11	Eesti	N	171	57	56	39	0	4	võlleib	jah	jah	jah
12	Soome	N	175	68	56	40	4	4	ei söö teav	jah	jah	jah
13	Eesti	N	168	70	57	40						
14	Eesti	N	152	50	52	37						
15	Eesti	N	164	55	54	37						
16	Eesti	N	160	60	55	37						
17	Eesti	N	163,3	67	56	39						
18	Eesti	N	163	59	56	37						
19	Soome	N	162	62	55	37						
20	Soome	N	158	65	56	37						
21	Eesti	N	170	57	57	39						
22	Eesti	N	170	58	56	37,5						
23	Eesti	N	162	57	52	38						
24	Eesti	N	169	63	56	39						
25	Eesti	N	176	67	54	41						
26	Eesti	N	173	64	54,5	41						
27	Eesti	N	170	63	54	38						
28	Eesti	N	167	73	56	39						
29	Eesti	N	167	58	56	39						
30	Eesti	M	178	80	59	42						
31	Eesti	M	189	71	60	47						
32	Eesti	M	181	61	55	46						
33	Eesti	M	189	71	60	47						
34	N	166	64	57	39							
35	Soome	N	169	65	57	39						
36	Soome	N	180	65	58	38						
37	Soome	N	183	59	41							
38	Soome	N	175	60	56	39						
39	Eesti	N	178	77	57	39						
40	Eesti	M	175	64	52	39						
41	Eesti	M	178	65	57	42						
42	Eesti	N	168	52	50	37,5	1	4	võlleib	jah	jah	jah
43	N	160	62	53	36	2	4	võlleib	jah	jah	jah	jah
44	Eesti	M	185	80	57	45	3	5	muu	jah	ei	ei
45	Eesti	N	175	85	59	41	3	5	võlleib	jah	jah	jah
46	Eesti	N	167	65	57	39	0	5	võlleib	nii ja nee	jah	ei
47	Eesti	N	176	74	57	43	3	4	võlleib	ei	jah	ei
48	Eesti	N	161	62	56	38	2	4	puder	jah	jah	ei
49	Soome	N	163	52	55	36	2	4	puder	jah	jah	ei
50	Eesti	N	161	62	56	38	2	4	puder	jah	jah	ei
51	Eesti	N	161,5	55	54	38	2	4	võlleib	nii ja nee	jah	jah
52	Eesti	N	163	59	53	37	2	4	võlleib	nii ja nee	jah	jah
53	Eesti	N	172	65	56	39	2	5	võlleib	nii ja nee	jah	ei
54	Eesti	M	184	83	58	44	1	5	võlleib	jah	jah	jah
55	Eesti	N	164	75	60	39	0	4	muu	nii ja nee	jah	jah
56	Eesti	N	168	57	56	36	5	4	võlleib	nii ja nee	jah	ei
57	Eesti	N	168	60	56	38	2	3	puder	jah	jah	ei
58	Eesti	N	160	49	52	37	2	4	heibed või	jah	jah	jah
59												
60	Vaatluste arv			=COUNT(C2:C58)								
61	Keskmine											
62	Mediaan											
63	Standardhälve											
64	Standardviga											
65	Min											
66	Max											

**Function Arguments**

COUNT

Value1: C2:C58 = {170;161;183;183;174;171;161;172...}

Value2: = number

= 57

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 = 57

[Help on this function](#)

OK Cancel

Vaatluste arv	57
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 '='!)

Vaatluste arv	=COUNT(C2:C58)
Keskmine	=AVERAGE(C2:C58)
Mediaan	=MEDIAN(C2:C58)
Standardhälve	=STDEV(C2:C58)
Standardviga	
Min	=MIN(C2:C58)
Max	=MAX(C2:C58)

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

- 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
59				
60	Vaatluste arv		57	
61	Keskmine		170,5	
62	Mediaan		170	
63	Standardhälve		8,5	
64	Standardviga		=C63/SQRT(C60)	
65	Min		152	
66	Max		189	

3. Rakendage samu funktsioone ning arvutage soovitud arvarakteristikute väärtused kõigi andmestikus sisalduvate arvtunnuste jaoks (va õlle tarbimine).

57	Eesti	N	168	60	56	38	2	3 puc
58	Eesti	N	160	49	52	37	2	4 hel
59								
60	Vaatluste arv		57					
61	Keskmine		170,5					
62	Mediaan		170					
63	Standardhälve		8,5					
64	Standardviga		1,1					
65	Min		152					
66	Max		189					

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

Tulemus:

Vaatluste arv	57	56	57	57	57	57
Keskmine	170,5	64,9	55,8	39,5	1,9	4,0
Mediaan	170	64	56	39	2	4
Standardhälve	8,5	8,9	2,4	2,7	1,3	0,7
Standardviga	1,1	1,2	0,3	0,4	0,2	0,1
Min	152	49	48	36	0	1
Max	189	85	60	47	8	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
2	Eesti	N	170	63	=D2/((C2/100)^2)	
3	Eesti	N	161	57		55

**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.

**Pane tähele** – mitteteadaoleva kehamassiga tudengi kehamassiindeksina väljastab *Excel* ilmselgelt ebaloomuliku väärtuse 0!!

	B	C	D	E	F
1	SUGU	PIKKUS	MASS	KMI	PEA_P
35	N	169	65	22,7583	57
36	N	180	65	20,0617	58
37	N	183		0	59
38	N	175	60	19,5918	56
39	N	178	77	24,3025	57

Lihtsaim viis taolise väärtuse korrigeerimiseks on see lihtsalt ära kustutada (sest me ei tea õiget kehamassi arvutamaks korrektset kehamassiindeksit).

	B	C	D	E	F
1	SUGU	PIKKUS	MASS	KMI	PEA_P
35	N	169	65	22,7583	57
36	N	180	65	20,0617	58
37	N	183		Delete	59
38	N	175	60	19,5918	56
39	N	178	77	24,3025	57

6. Leidke vajalikud arvkarakteristikute väärtused ka uuele tunnusele.

Eesti	N	160		52	3
Copy -> Paste					
Vaatluste arv		57	56	56	5
Keskmine		170,5	64,9	22,4	39,
Mediaan		170	64	21,9806	3
Standardhälve		8,5	8,9	2,3	2,
Standardviga		1,1	1,2	0,3	0,
Min		152	49	18,424	3
Max		189	85	27,8852	4

	A	B	C	D	E	F	G	H
1	RIIK	SUGU	PIKKUS	MASS	KMI	PEA_P	JALANR	ODE
2	Eesti	N	170	63	21,7583	37	39	
3	Eesti	N	161	57		39	37,3	
4	Eesti	M	183	80		48	43	
5	Eesti	N	183	81		38	42	
6	Eesti	M	174	74		36	43	
7	Eesti	N	171	70		37	39	
8	Eesti	N	161	58		35	37	
9	Eesti	N	172	80		36	41	
10	Eesti	N	183,3	68		33,3	40	
11	Eesti	N	171	57		36	39	
12	Soome	N	175	68		36	40	
13	Eesti	N	168	70		37	40	
14	Eesti	N	152	50		32	37	
15	Eesti	N	164	35		34	37	
16	Eesti	N	160	60		35	37	
17	Eesti	N	163,3	67		36	39	
18	Eesti	N	163	33		36	37	
19	Soome	N	162	62		35	37	
20	Soome	N	158	65		36	37	
21	Eesti	N	170	57		37	39	
22	Eesti	N	170	38		36	37,3	
23	Eesti	N	162	57		32	38	
24	Eesti	N	169	63		36	39	
25	Eesti	N	176	67		34	41	
26	Eesti	N	173	64		34,3	41	
27	Eesti	N	170	63		34	38	
28	Eesti	N	167	73		36	39	
29	Eesti	N	167	38		36	39	
30	Eesti	M	178	80		39	42	
31	Eesti	M	189	71		60	47	
32	Eesti	M	181	61		33	46	
33	Eesti	M	189	71		60	47	
34	N		166	64		37	39	
35	Soome	N	169	65		37	39	
36	Soome	N	180	65		38	38	
37	Soome	N	183			38	41	
38	Soome	N	175	60		36	39	
39	Eesti	N	178	77		37	39	
40	Eesti	M	173	64		32	39	
41	Eesti	M	178	65		37	42	
42	Eesti	N	168	52		30	37,3	
43	N		160	62		33	36	
44	Eesti	M	185	80		37	45	
45	Eesti	N	175	85		38	41	
46	Eesti	N	167	65		33	39	
47	Eesti	N	176	74		37	43	
48	Eesti	N	161	62		36	38	
49	Soome	N	163	52		33	36	
50	Eesti	N	161	62		36	38	
51	Eesti	N	161,3	35		34	38	
52	Eesti	N	163	33		33	37	
53	Eesti	N	172	65		36	39	
54	Eesti	M	184	83		38	44	
55	Eesti	N	164	75		60	39	
56	Eesti	N	168	57		36	36	
57	Eesti	N	168	60		36	38	
58	Eesti	N	160	48		32	37	

## 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*

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* kaheteistkümmne 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 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:

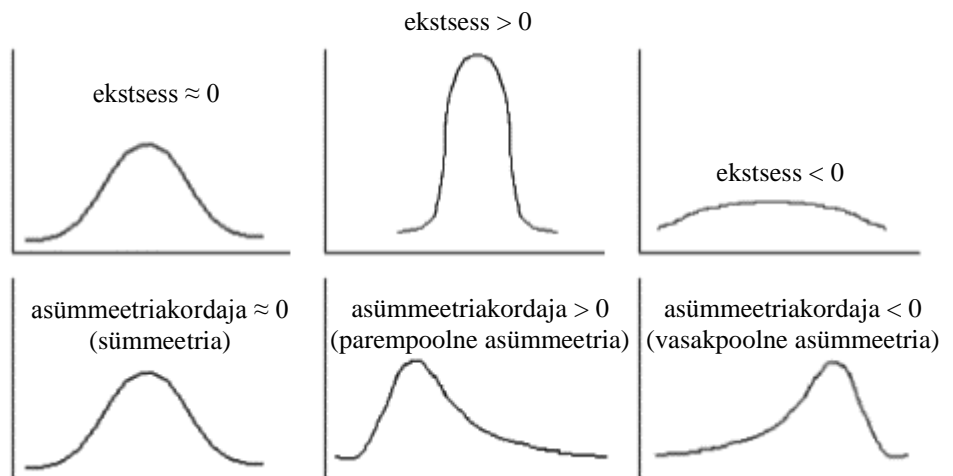
PIKKUS	MASS	KMI	PEA_P	JALANR							
Mean	170,461	Mean	64,875	Mean	22,352	Mean	55,7895	Mean	39,5175	} Valiku <i>Summary statistics</i> tulemus	Keskmine
Standard Error	1,12742	Standard Error	1,19525	Standard Error	0,3127	Standard Error	0,31162	Standard Error	0,35288		Standardviga
Median	170	Median	64	Median	21,9806	Median	56	Median	39		Mediaan
Mode	170	Mode	65	Mode	21,7993	Mode	56	Mode	39		Mood
Standard Deviation	8,51181	Standard Deviation	8,9444	Standard Deviation	2,34	Standard Deviation	2,35271	Standard Deviation	2,66419		Standardhälve
Sample Variance	72,451	Sample Variance	80,0023	Sample Variance	5,47561	Sample Variance	5,53524	Sample Variance	7,0979		Dispersioon
Kurtosis	-0,54341	Kurtosis	-0,45558	Kurtosis	-0,46323	Kurtosis	1,5411	Kurtosis	1,10543		Ekstsess e järsakuskordaja
Skewness	0,3337	Skewness	0,45246	Skewness	0,44651	Skewness	-0,78045	Skewness	1,2026		Asümmeetriakordaja
Range	37	Range	36	Range	9,46115	Range	12	Range	11		Ulatus = Max - Min
Minimum	152	Minimum	49	Minimum	18,424	Minimum	48	Minimum	36		
Maximum	189	Maximum	85	Maximum	27,8852	Maximum	60	Maximum	47		
Sum	9716,3	Sum	3633	Sum	1251,71	Sum	3180	Sum	2252,5		
Count	57	Count	56	Count	56	Count	57	Count	57		Vaatluste arv
Largest(2)	189	Largest(2)	83	Largest(2)	27,7551	Largest(2)	60	Largest(2)	47		
Smallest(2)	158	Smallest(2)	50	Smallest(2)	18,6197	Smallest(2)	50	Smallest(2)	36		
Confidence Level(95,0%)	2,25849	Confidence Level(95,0%)	2,39533	Confidence Level(95,0%)	0,62666	Confidence Level(95,0%)	0,62426	Confidence Level(95,0%)	0,7069		

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

Enamusest protseduuri *Descriptive Statistics* väljundis sisalduvatest arvarakteristikute 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 märkimisväärsest erinevusest normaalkujust on mõtet rääkida siis, kui ükskõik kumb neist kordajatest omandab absoluutväärtuselt 1-st suurema väärtuse ...

Eriti palju neid kordajaid siiski ei kasutata.



- **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 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	170,461
4	Standard Error	1,12742
5	Median	170
6	Mode	170
7	Standard Deviation	8,51181
8	Sample Variance	72,451
9	Kurtosis	-0,54341
10	Skewness	0,3337
11	Range	37
12	Minimum	152
13	Maximum	189
14	Sum	9716,3
15	Count	57
16	Largest(2)	189
17	Smallest(2)	158
18	Confidence Level(95,0%)	2,25849
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	168,209
Ülemine 95% usalduspiir	172,7199

Seega, tõlgendades antud andmestikku kui valimit Maaülikooli veterinaarmeditsiini eriala esmakursuslastest aastal 2010, võib väita, et esimese kursuse tudengite keskmine pikkus jääb 95% tõenäosusega vahemikku 168,2 cm kuni 172,7 cm. St, et mõõtes ära **kõigi** teie kursuse tudengite (ka nende, kelle andmeid mingil põhjusel selles andmetabelis pole) 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 (168,6; 172,3) ja (167,5; 173,5). **Miks on 90% usaldusintervall kitsam?**
- **Arvutage usaldusintervall ka teiste tunnuste keskmistele väärtustele ning püüdke neist vähemalt ühe kohta sõnastada lõppjärelus!!**