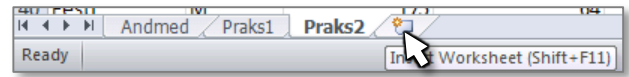


## Biomeetria praks 3

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

1. Avage *MS Excel*'is 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	D
1	GENDER	HEIGHT	WEIGHT	HEIGHT
54	W	162	70	
55	W	172	58	
56				
57	Vaatluste arv			
58	Keskmine			
59	Mediaan			
60	Standardhälve			
61	Standardviga			
62	Min			
63	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):

The image illustrates the steps to use the COUNT function in Excel. It shows a spreadsheet with student data, the 'Insert Function' dialog box where 'COUNT' is selected from the 'Statistical' category, and the 'Function Arguments' dialog box where the range 'B2:B55' is entered. The final result of 54 is shown in a summary table.

1	A	B	C	HE
54	W	162	70	
55	W	172	58	
56				
57	Vaatluste arv			
58	Keskmine			

A	B	C	D	E	F	G	H		
5	W	170	50	33	37	4	sandwich	yes	ye
6	W	178	68	38	41	3	cereals or	yes	ye
7	W	163	56	40	37	4	sandwich	yes	ye
8	W	177	63	33	40	3	sandwich	sometime	ye
9	W	162.5	53	33	38	3	porridge	yes	ye
10	W	170	73	36	39	3	other	yes	ye
11	M	173	74	37	42	3	sandwich	yes	ye
12	W	176	66	37	39	4	sandwich	sometime	no
13	M	173	64	36	42	4	other	yes	ye
14	M	190	82	38	46	4	other	yes	ye
15	W	161	50	33	37	4	nothing	no	ye
16	W	170	83	37	41	4	cereals or	no	ve
17	W	176	38	32	39				
18	W	172	90	38	41				
19	W	138	33	37	38				
20	M	189	82	17	43				
21	W	169	60	33.3	41				
22	W	164	32	36	37				
23	W	172	62	36	39				
24	W	173	66	36	40				
25	W	169	60	33	39				
26	W	162	50	30	38				
27	W	163	32	30.3	37				
28	M	170	80	36	41				
29	M	176	74	36	42				
30	M	173	73	34	43				
31	W	171	63	37	39				
32	W	170	60	33	39				
33	W	163	62	33	38				
34	M	181	74	33	44				
35	W	168	60	33	39				
36	W	174	34	33	40				
37	W	166	68	36	39				
38	W	168	63	33	39				
39	W	163	58	36	37				
40	W	171	73	33	41				
41	W	163	77	39	39				
42	W	161	33	27	38				
43	M	183	73	33	43				
44	W	169	33	33	38				
45	W	173	60	37	42				
46	W	167	80	37.3	41				
47	W	138	70	33	38				
48	M	174	87	37	40				
49	W	163	61	37	39				
50	W	164	38	37	39				
51	W	183	80	60	41				
52	W	177	63	60	40				
53	W	160	70	37	39				
54	W	162	70	33	40				
55	W	172	58	62	39				
56									
57	Vaatluste arv	=COUNT(B2:B55)							
58	Keskmine								

Vaatluste arv	54
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(B2:B55)
Keskmine	=AVERAGE(B2:B55)
Mediaan	=MEDIAN(B2:B55)
Standardhälve	=STDEV.S(B2:B55)
Standardviga	
Min	=MIN(B2:B55)
Max	=MAX(B2:B55)

=B60/SQRT(B57)			
	A	B	C
56			
57	Vaatluste arv	54	
58	Keskmine	170.1944	
59	Mediaan	170	
60	Standardhälve	7.397358	
61	Standardviga	=B60/SQRT(B57)	
62	Min	158	
63	Max	190	

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

1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZE	MATH	BR
54	W	162	70	55	40		5 sar
55	W	172	58	62	39		4 ot
56							
57	Vaatluste arv	54					
58	Keskmine	170.1944					
59	Mediaan	170					
60	Standardhälve	7.397358					
61	Standardviga	1.006653					
62	Min	158					
63	Max	190					

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

Tulemus:

	A	B	C	D	E	F	BR
1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZE	MATH	BR
56							
57	Vaatluste arv	54	54	51	54	54	
58	Keskmine	170.2	65.4	55.9	39.6	4.0	
59	Mediaan	170	63	56	39	4	
60	Standardhälve	7.4	10.7	2.1	2.0	0.8	
61	Standardviga	1.0	1.5	0.3	0.3	0.1	
62	Min	158	47.5	50	36	2	
63	Max	190	90	62	46	5	

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

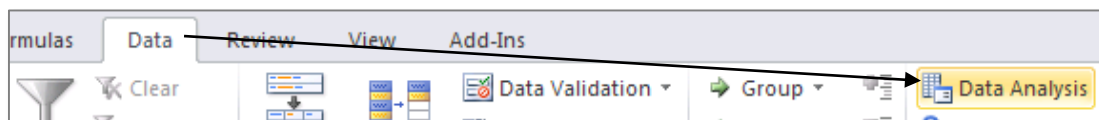


## Ülesanne 2.

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

### Tööjuhend

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



1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	GENDER	HEIGHT	WEIGHT	BMI	HEAD	SHOE_SIZE	MATH	BREAKFAST	PORRIDGE	PET	SICK	SPORT	SKI	CAR	BEER	SMOKE		
3	W	170	70	24.2215	55.3	39	3	other	yes	yes	no	yes	yes	yes	0	no		
4	W	158	47.3	19.0274	53	36	3		yes	yes	no	yes	yes	yes	0	no		
5	W	170	60	20.7612	53	38	3		yes	yes	no	yes	yes	yes	0	no		
6	W	170	30	17.301	53	37	3		yes	yes	no	yes	yes	yes	0	no		
7	W	179	68	21.2228	58	41	3		yes	yes	no	yes	yes	yes	0	no		
8	W	163	56	21.0772	40	37	3		yes	yes	no	yes	yes	yes	0	no		
9	W	177	63	20.7476	53	40	3		yes	yes	no	yes	yes	yes	0	no		
10	W	162.3	53	20.071	53	38	3		yes	yes	no	yes	yes	yes	0	no		
11	W	170	75	25.9516	56	39	3		yes	yes	no	yes	yes	yes	0	no		
12	M	173	74	24.1633	57	42	3		yes	yes	no	yes	yes	yes	0	no		
13	W	176	66	21.3068	57	39	3		yes	yes	no	yes	yes	yes	0	no		
14	M	173	64	20.898	56	42	3		yes	yes	no	yes	yes	yes	0	no		
15	M	190	82	22.7147	58	46	3		yes	yes	no	yes	yes	yes	0	no		
16	W																	
17	W																	
18	W																	
19	W																	
20	M																	
21	W	169	60	21.0077	55.3	41	3		yes	yes	no	yes	yes	yes	0	no		
22	W	164	52	19.3337	56	37	3		yes	yes	no	yes	yes	yes	0	no		
23	W	172	62	20.9573	56	39	3		yes	yes	no	yes	yes	yes	0	no		
24	W	173	66	22.0322	56	40	3		yes	yes	no	yes	yes	yes	0	no		
25	W	169	60	21.0077	53	39	3		yes	yes	no	yes	yes	yes	0	no		
26	W	162	50	19.052	50	38	3		yes	yes	no	yes	yes	yes	0	no		
27	W	165	52	19.1001	50.5	37	3		yes	yes	no	yes	yes	yes	0	no		
28	M	170	80	27.701	55	41	3		yes	yes	no	yes	yes	yes	0	no		
29	M	176	74	23.883	55	41	3		yes	yes	no	yes	yes	yes	0	no		
30	M	173	73	24.303	55	41	3		yes	yes	no	yes	yes	yes	0	no		
31	W	171	63	21.3068	53	39	3		yes	yes	no	yes	yes	yes	0	no		
32	W	170	60	20.9017	53	39	3		yes	yes	no	yes	yes	yes	0	no		
33	W	163	62	23.3353	53	38	3		yes	yes	no	yes	yes	yes	0	no		
34	M	181	74	22.5878	53	44	3		yes	yes	no	yes	yes	yes	0	no		
35	W	168	60	21.2383	53	39	3		yes	yes	no	yes	yes	yes	0	no		
36	W	174	54	17.8359	53	40	3		yes	yes	no	yes	yes	yes	0	no		
37	W	166	68	24.677	56	39	3		yes	yes	no	yes	yes	yes	0	no		
38	W	168	63	22.3214	53	39	3	sandwich	yes	yes	yes	yes	yes	yes	0.5	yes		
39	W	163	58	21.3039	56	37	3	sandwich	no	yes	no	yes	yes	no	1	no		
40	W	171	75	25.6489	53	41	3	sandwich	yes	yes	no	yes	yes	yes	0	no		
41	W	163	77	28.2828	58	39	3	sandwich	yes	yes	no	no	no	yes	0	no		
42	W	161	55	21.2183	57	38	3	porridge	yes	yes	yes	yes	yes	yes	0	no anymore, but I've smoked		
43	M	183	75	22.3954	73	43	3	porridge	yes	no	no	yes	yes	yes	3	no		
44	W	169	53	18.5368	53	38	3	sandwich	sometime	yes	no	yes	yes	yes	0	no		
45	W	173	60	19.5918	57	42	3	cereals or muesli	yes	yes	no	no	no	no	0.5	no		
46	W	167	80	28.6851	57.5	41	3	other	yes	yes	no	yes	yes	yes	2	no		
47	W	158	70	28.0404	53	38	3	cereals or muesli	yes	yes	yes	yes	yes	no	0	no		
48	M	174	87	28.7356	57	40	4	sandwich	sometime	yes	yes	yes	no	yes	0.5	no		
49	W	165	61	22.4059	57	39	3	other	sometime	yes	yes	no	no	yes	0.5	yes		
50	W	164	58	21.3643	57	39	3	sandwich	yes	yes	yes	yes	yes	yes	0	no anymore, but I've smoked		
51	W	185	80	23.3747	60	41	4	cereals or muesli	sometime	yes	no	yes	no	yes	0	no		
52	W	177	63	20.1092	60	40	2	sandwich	no	no	no	yes	yes	no	0	no		
53	W	160	70	27.3438	57	39	4	sandwich	sometime	yes	yes	yes	no	yes	0	no anymore, but I've smoked		
54	W	162	70	26.6728	53	40	3	sandwich	no	yes	no	no	no	no	2	no		
55	W	172	58	19.6052	62	39	4	other	sometime	yes	no	yes	no	no	0	no anymore, but I've smoked		

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* kaheteistkümne põhilise arvkarakteristiku 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:

HEIGHT	WEIGHT	BMI	HEAD	SHOE_SIZE							
Mean	170.19	Mean	65.435	Mean	22.557	Mean	55.863	Mean	39.648	} Valiku <i>Summary statistics</i> tulemus	Keskmine
Standard Error	1.0067	Standard Error	1.4509	Standard Error	0.4386	Standard Error	0.2964	Standard Error	0.2736		Standardviga
Median	170	Median	63	Median	21.555	Median	56	Median	39		Mediaan
Mode	170	Mode	60	Mode	20.761	Mode	55	Mode	39		Mood
Standard Deviation	7.3974	Standard Deviation	10.662	Standard Deviation	3.2227	Standard Deviation	2.1168	Standard Deviation	2.0109		Standardhälve
Sample Variance	54.721	Sample Variance	113.68	Sample Variance	10.386	Sample Variance	4.4808	Sample Variance	4.0437		Dispersioon
Kurtosis	0.3472	Kurtosis	-0.6752	Kurtosis	-0.1694	Kurtosis	1.94	Kurtosis	0.7617		Ekstsess e järsakuskordaja
Skewness	0.5808	Skewness	0.3989	Skewness	0.7812	Skewness	-0.1468	Skewness	0.7797		Asümmeetriakordaja
Range	32	Range	42.5	Range	13.121	Range	12	Range	10		Ulatus = Max - Min
Minimum	158	Minimum	47.5	Minimum	17.301	Minimum	50	Minimum	36		
Maximum	190	Maximum	90	Maximum	30.422	Maximum	62	Maximum	46		
Sum	9190.5	Sum	3533.5	Sum	1218.1	Sum	2849	Sum	2141		
Count	54	Count	54	Count	54	Count	51	Count	54		Vaatluste arv
Largest(2)	189	Largest(2)	87	Largest(2)	29.412	Largest(2)	60	Largest(2)	44		
Smallest(2)	158	Smallest(2)	50	Smallest(2)	17.836	Smallest(2)	50.5	Smallest(2)	37		
Confidence Level(95.0%)	2.0191	Confidence Level(95.0%)	2.9102	Confidence Level(95.0%)	0.8796	Confidence Level(95.0%)	0.5954	Confidence Level(95.0%)	0.5489		

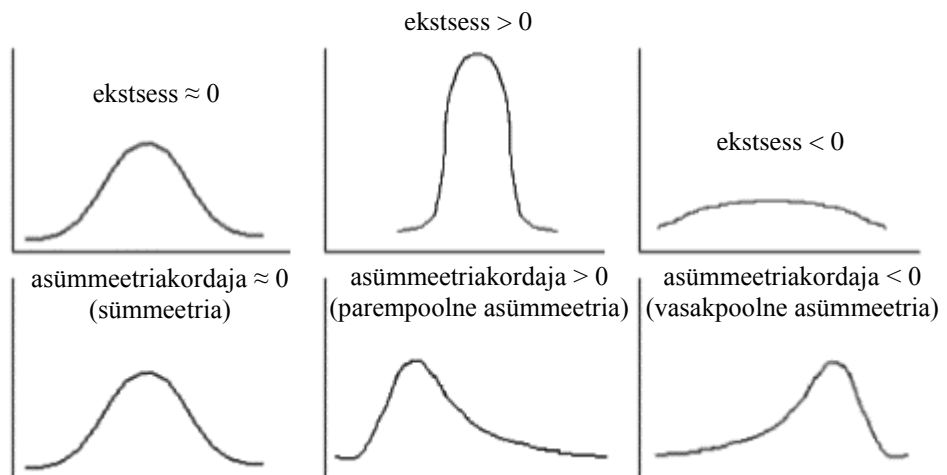
• **Lisalugemine – uuritava tunnuse jaotuse kuju iseloomustamine**

Enamusest protseduuri *Descriptive Statistics* väljundis sisalduvatest arvkarakteristikuteist 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 peaümbermõõdule vastav järsakuskordaja (1,94) pisut suurem võrreldes teiste kehamõõtude sama parameetriga – see näitab, et enamuste tudengite peaümbermõõdu väärtused jäävad küllalt kitsasse vahemikku, samas leiduvad üksikud teistest palju suurema või väiksema peaümbermõõdu 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 eneselt leida.

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

	R	S
1	HEIGHT	
2		
3	Mean	170.19
4	Standard Error	1.0067
5	Median	170
6	Mode	170
7	Standard Deviation	7.3974
8	Sample Variance	54.721
9	Kurtosis	0.3472
10	Skewness	0.5808
11	Range	32
12	Minimum	158
13	Maximum	190
14	Sum	9190.5
15	Count	54
16	Largest(2)	189
17	Smallest(2)	158
18	Confidence Level(95.0%)	2.0191
19		
20	Alumine 95% usalduspiir	=S3-S18
21	Ülemine 95% usalduspiir	=S3+S18

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	168.18
Ülemine 95% usalduspiir	172.21

Seega, tõlgendades antud andmestikku kui valimit Maaülikooli esmakursuslastest õppeaastal 2013-2014, võib väita, et esimese kursuse tudengite **keskmine pikkus** jääb 95%-lise tõenäosusega vahemikku 168,2 cm kuni 172,2 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 (168,5; 171,9) ja (167,5; 172,9). **Miks on 90%-line usaldusintervall kitsam?**
- **Arvutage usaldusintervall ka teiste tunnuste keskmisele väärtustele ning püüdke neist vähemalt ühe kohta sõnastada lõppjärelus!!**