

Biomeetria praks 2

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excelis* oma kursuse ankeedivastuseid sisaldav andmestik (see, mida 1. praktikumiski analüüsisite),
2. nimetage 'Sheet3' ümber leheküljeks 'Praks2' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel (ehk algandmed) lehekülje 'Praks2' ülemisse vasakusse nurka.

Ülesanne 1.

Konstrueerige sagedustabel tunnusele 'PEA_P' (peaübermõõt) ja illustreerige saadud tabelit histogrammiga. Kujundage histogramm.

Tööjuhend

Pidevale arvtunnusele sagedustabeli tegemiseks on *MS Excelis* kaks moodust – statistika-protseduur *Histogram* ja funktsioon *FREQUENCY*. Mõlemad need variandid eeldavad, et kasutaja on eelnevalt välja mõelnud loodavad klassid ja sisestanud klasside ülemised piirid *Exceli* töölehele

(tegelikult võib protseduur *Histogram* moodustada klassid ka ise, aga siis on nende piirid sageli „mitte ümmargused“ ja uuritava tunnuse väärtuste paiknemise kirjeldamine on seeläbi tülikam).

Järgnevalt tutvume lähemalt sagedustabeli konstrueerimisega statistikaprotseduuri *Histogram* abil (kuigi 1. punktis kirjeldatav klasside moodustamine ja nende *Excelile* ette andmine toimub analoogselt ka funktsiooni *FREQUENCY* korral).

1. Otsustamaks klasside arvu ja suuruse üle, peab esmalt omama ülevaadet vaatluste arvust ja uuritava tunnuse väärtuste ulatusest.
 - Vaatluste arv $n = 52$ (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaübermõõtude väärtused võiks jagada 7 või 8 klassi (sest $\sqrt{52} \approx 7,2$; samas ei ole see absoluutne tõde, kui klasside piiridega paremini sobib, võib väärtused jagada ka 6 või hoopis 9 klassi).
 - Väärtuste paiknemise ulatuse määravad ära minimaalne ja maksimaalne väärtus. Nende leidmiseks on *Excel'*is päris palju erinevaid mooduseid, kiireim variant on järgmine (vt ka joonist järgmisel leheküljel):
 - a. võtate blokki huvipakkuvad lahtrid,
 - b. klikite hiire parempoolse klahviga *Exceli* tööakna alumisel ribal (nn olekuribal, vt joonist järgmisel lehel) ja märgite avanenud rippmenüüst teid huvitavad funktsioonid (nt. *Min*),
 - c. valitud funktsioonide väärtused kuvatakse samas akna alumisel serval, aga seda ei trükita kuhugi tabeli lahtrisse.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VENI	MAT_HINI	HOMMIK	PUDER	LEMMIK	HAIGE	SPORT
2	Eesti	M	188	88	56	42	2	5	võleib	jah	jah	ei	jah
3	Eesti	N	167	65	55	39	1	5	puder	jah	jah	ei	jah
4	Eesti	N	179	68	54	42	2	4	võleib	jah	ei	ei	jah
5	Eesti	N	178	60	55	38	2	4	puder	nii ja naa	jah	ei	jah
6	Eesti	N	161	70	56,5	39	1	3	võleib	jah	ei	ei	ei
7	Eesti	M	185	69	56	44	1	4	muu	jah	jah	ei	jah
8	Soome	N	174	58	58	40	10	4	puder	jah	jah	ei	jah
9	Eesti	N	171	55	41	39	2	4	võleib	jah	jah	ei	jah
10	Eesti	N	165	55	42	35	0	3	puder	nii ja naa	jah	ei	jah
11	Soome	N	167	65	58,5	38	1	5	puder	nii ja naa	jah	ei	ei
12	Soome	N	163	69	58	38	2	5	muu	jah	jah	ei	jah
13	Soome	N	165	55	58	38	1	4	võleib	nii ja naa	jah	ei	jah
14	Soome	N	169	61	58	36	2	4	muu	jah	jah	ei	jah
15	Soome	N	170	68	58	38	3	4	võleib		jah	jah	jah
16	Eesti	N	165	51	56	39	1	4	helbed või muu	nii ja naa	jah	ei	jah
17	Eesti	N	170	57	56								
18	Eesti	N	170	63	56								
19	Soome	N	162	65	58								
20	Eesti	N	168	80	56								
21	Eesti	N	177	60									
22	Eesti	N	168	73	53								
23	Eesti	N	165	50,1	52								
24	Eesti	M	188	80	58								
25	Eesti	N	162	63	56								
26	Soome	N	162	52	55								
27	Eesti	M	183	73	54								
28	Soome	N	161	47	62								
29			163	53	58								
30	Soome	N	179	80	62								
31	Soome	N	179	68	57								
32	Soome	N	176	80	58								
33	Soome	N	155	52	53								
34	Eesti	N	162	57	53								
35	Eesti	M	190	85	58								
36	Eesti	M	184	78	55								
37	Eesti	N	167	66	54								
38	Eesti	M	175	62	56,5								
39	Eesti	N	180	71	56								
40	Eesti	N	164	52	56								
41	Eesti	M	183	80	56								
42	Eesti	N	179	75	56								
43	Venemaa	N	167	46,5	58								
44	Eesti	N	172	64	55								
45	Soome	N	167	55	50								
46	Soome	N	171	75	56								
47	Soome	N	160	61	54								
48	Soome	N	168	75	58								
49	Soome	N	163	50	54								
50		N	169	65	57								
51	Soome	N	172	54	55								
52	Eesti	N	163	52	48								
53	Eesti	N	164	55	55,5								
54	Soome	N	156	48	50								
55													

Customize Status Bar

- Cell Mode Ready
- Signatures Off
- Information Management Policy Off
- Permissions Off
- Caps Lock Off
- Num Lock Off
- Scroll Lock Off
- Fixed Decimal Off
- Overtyping Mode
- End Mode
- Macro Recording Not Recording
- Selection Mode
- Page Number
- Average 55,17307692
- Count 52
- Numerical Count 52
- Minimum 41
- Maximum 62
- Sum 2869
- Upload Status
- View Shortcuts
- Zoom 70%
- Zoom Slider

- Maksimaalne peaümberrõõm tuleb 62 ja minimaalne 41 cm. Kas need arvud on reaalsed? Kui ümberrõõm on omale keeruline ette kujutada, siis diameetriga (läbimõõduga) on ehk lihtsam. Siit ka küsimus – kui suur on pea läbimõõm, kui ümberrõõm on 41 cm, ja kas see on reaalne?

Spikker. Mäletatavasti avaldub ringi ümberrõõm valemiga $P = 2\pi r$, millest diameeter $d = 2r = P/\pi$ (Excelit arvutusmasina kasutades võite sisestada valemi kujul '=41/PI()').

Tulemuseks peaks tulema 13,1 cm. Võrdluseks – tavalise CD-plaadi läbimõõm on 12 cm. Seega tundub peaümberrõõm 41 cm pisut liiga väike.

Ilmselt on tegu mõõtmis- või andmete ülesmärkimisveaga.

Et traditsioonilised statistikameetodid ei ole mõeldud ebareaalsete või väga erandlike andmete analüüsiks, jäetakse sellised väärtused sageli analüüsist välja. Antud juhul on kõige mõistlikum peaümberrõõm väärtus 41 cm andmetabelist lihtsalt ära kustutada.

Peale peaümbermõõdu väärtuse 41 cm kustutamist selgub, et andmebaasis leidub ka peaümbermõõd 42 cm, seegi on pisut liiga väike (ilmselt on mõõt võetud pisut valest kohast), mistap on objektiivsemate tulemuste huvides mõistlik seegi väärtus andmebaasist kustutada.

Ja ära kustutada tuleks need väärtused nii 'Praks2'-lehelt kui ka edasiste segaduste vältimiseks 'Andmed'-lehelt.

58	58	40
55	41	39
55	42	35
RR	RR R	RR

→

58	58	40
55		39
55		35
RR	RR R	RR

- Minimaalne peaümbermõõd korrigeeritud andmetest on 48 ja maksimaalne 62 cm.
- Klasside intervallide ligikaudseks leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovitatavate klasside arvuga ehk $(62-48)/8 = 1,75$ cm. Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli vajadusel veidi suurendada või vähendada. Antud juhul võiks näiteks valida klasside suuruseks 2 cm.
- Klasside piiride paika panekul alustatakse esimesest klassist, mis määratakse nii, et see sisaldaks minimaalset väärtust, näiteks võiks esimene klass olla 47-49 cm.
- Kokkuvõttes võime tudengite peaümbermõõdudest moodustada kaheksa 2-sentimeetrist klassi:

47-49; 49-51; 51-53; 53-55; 55-57; 57-59; 59-61; 61-63.

Et *Excel* paigutab klassi piiriga võrdsed väärtused alumisse klassi, on korrektsem esitada klassid poollõikudena (sest kuidas muidu aru saada, kumba klassi näiteks 51 cm kuulub?):

(47,49], (49,51], (51,53], (53,55], (55,57], (57,59], (59,61] ja (61,63].

NB. Soovi korral võinuks näiteks moodustada ka hoopis **kaheksa 2-sentimeetrist** klassi kujul

(46,48], (48,50], (50,52], (52,54], (54,56], (56,58], (58,60], (60,62];

või siis hoopis **viis 3-sentimeetrist** klassi kujul

(47,50], (50,53], (53,56], (56,59], (59,62].

- Moodustatud klasside *Excel*ile ette andmiseks tuleb sisestada klasside ülemisi piire sisaldav abitabel:

S	T	U
		pea_ymb klassid
0 päeva jooksul		49
uu jooksul		51
aasta jooksul		53
uu jooksul		55
0 päeva jooksul		57
11 aasta tagasi		59
11 aasta tagasi		61
11 aasta tagasi		

- Märkusi:
 - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;
 - *Excel* tõlgendab etteantud väärtusi järgmiselt:
 - '49' tähendab ' ≤ 49 ';
 - '51' tähendab ' $49 < x \leq 51$ ' jne
 (ehk igasse klassi kuuluvaks loetakse need väärtused, mis on väiksemad või võrdsed ette antud ülemisest klassi piirist ja mis ei kuulu eelnevatesse klassidesse);
 - viimast klassi '63'=(61,63] ette ei anta, sest sagedustabeli moodustamisel teeb *Excel* ise täiendava klassi, kuhu loeb kokku kõik eelnevatesse klassidesse mitte kuuluvad väärtused. Seega jäävad täiendavasse automaatselt loodavasse klassi kõik 61-st suuremad väärtused.

2. Sagedustabel protseduuri *Histogram* abil: *Data*-sakk → *Data Analysis...* → *Histogram*

The screenshot shows the Excel interface with the 'Data' tab selected. The 'Data Analysis' button in the ribbon is highlighted. A dialog box titled 'Data Analysis' is open, with 'Histogram' selected in the list of tools. A red dashed line traces the path from the 'Data' tab to the 'Data Analysis' button and then to the 'Histogram' option in the dialog box.

Andmed
Klassipiirid

Ütleb *Excel*'ile, et etteantud lahtri-
blokkide esimeses reas on nimed

Loodava tabeli (vasaku ülemise nurga)
asukoht

Lisaks tavalistele sagedustele
arvutatakse ka kumulatiivsed
suhtelised sagedused ehk jaotus

The 'Histogram' dialog box is shown with the following settings:

- Input Range:** \$E\$1:\$E\$54
- Bin Range:** \$U\$1:\$U\$8
- Labels**
- Output options:**
 - Output Range:** \$X\$1
 - New Worksheet Ply:**
 - New Workbook**
 - Pareto (sorted histogram)**
 - Cumulative Percentage**
 - Chart Output**

Protseduuri *Histogram* oodatav tulemus:

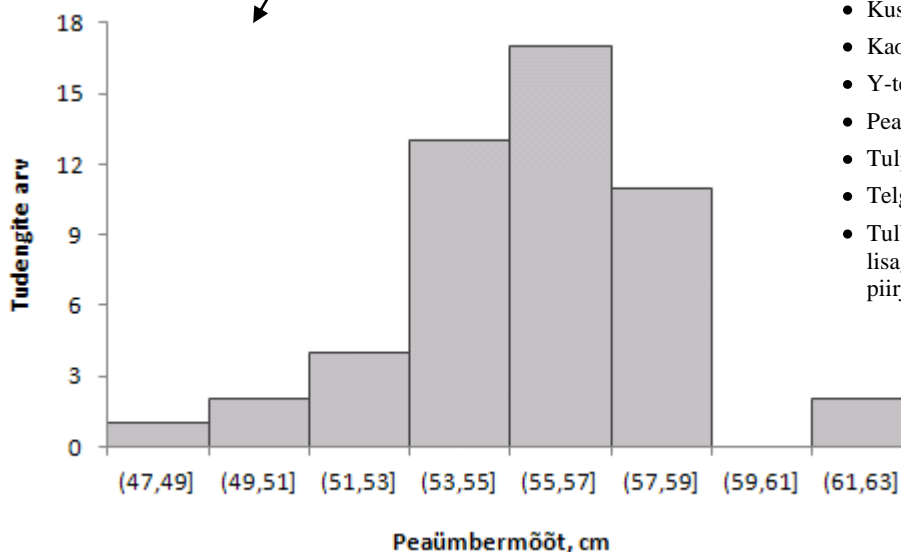
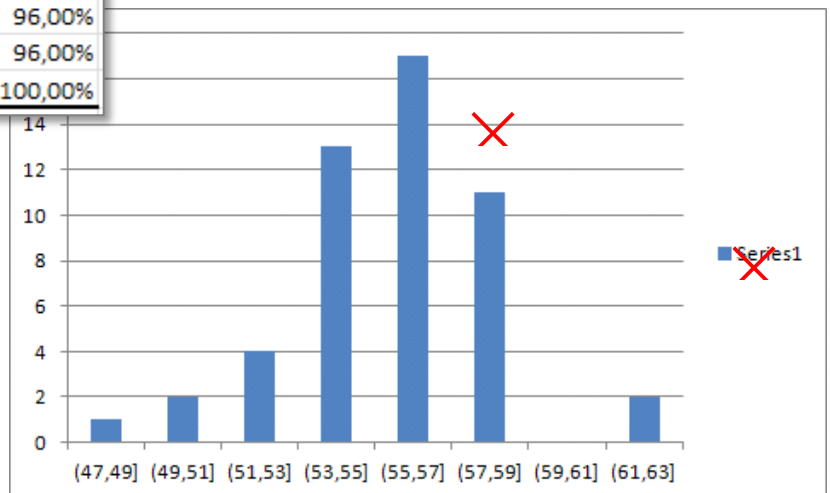
(kumulatiivne sagedus ehk **jaotus** näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

X	Y	Z
pea_ymb klassid	Frequency	Cumulative %
49	1	2,00%
51	2	6,00%
53	4	14,00%
55	13	40,00%
57	17	74,00%
59	11	96,00%
61	0	96,00%
More	2	100,00%

3. Lisage tabelile (näiteks tabeli ette tühja veergu) tegelikud klasside piirid ning tehke joonis.

	pea_ymb klassid	Frequency	Cumulative %
(47,49]	49	1	2,00%
(49,51]	51	2	6,00%
(51,53]	53	4	14,00%
(53,55]	55	13	40,00%
(55,57]	57	17	74,00%
(57,59]	59	11	96,00%
(59,61]	61	0	96,00%
(61,63]	More	2	100,00%

NB! Eraldi paiknevate lahtrite selekteerimiseks hoidke all 'Ctrl'-klahvi ...



- Kustutage legend ja ruudujooned
- Kaotage ära hall kast joonise ümbert
- Y-telje ulatuseks 0 kuni 18 sammuga 3
- Pealkirjad x- ja y-teljele
- Tulpade vahe = 0
- Telgede ühikud ja nimed kirjasuuruses 10
- Tulpad värvige nii, nagu enesele meeldib, lisage tulpadele ka sisust veidi tumedamad piirjooned

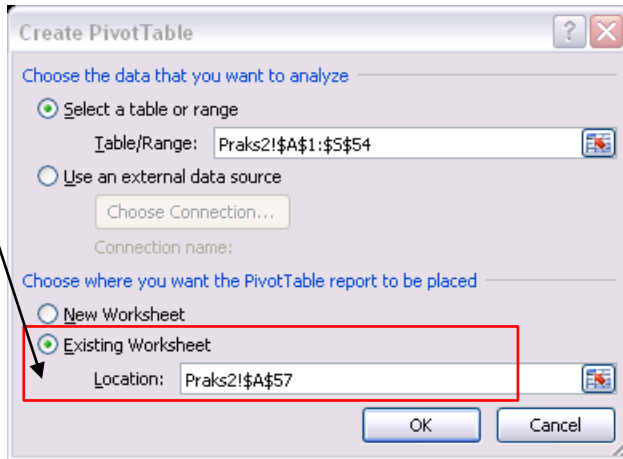
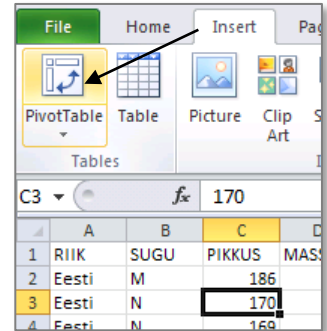
4. Sõnastage esimese ülesande lõpetuseks üks lause absoluutsete sageduste ja üks lause kumulatiivsete sageduste põhjal – pange need laused ka kirja.

Ülesanne 2.

Konstrueerige *Pivot Table*'i abil tabel, kirjeldamaks tudengite kehamassi sõltuvalt pudru söömisest.

Tööjuhend

1. Paigutage kursor andmetabeli suvalisse lahtrisse ('Praks2'-lehel).
2. *Insert*-sakk → *PivotTable*
3. Loodav tabel paigutage 'Praks2'-lehele andmetabeli alla:

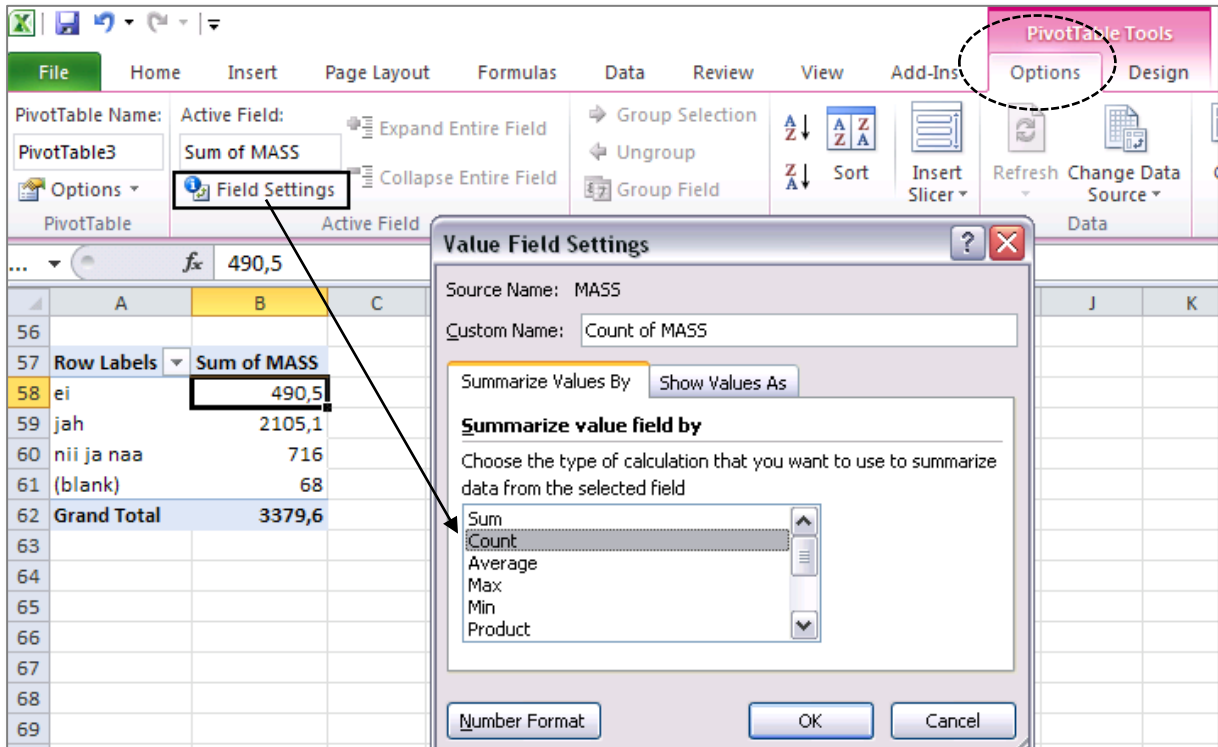


4.

Tulemuseks saadud tabel:

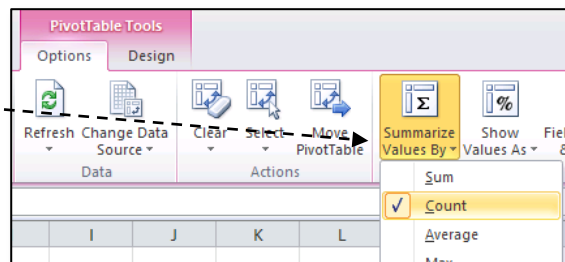
Row Labels	Sum of MASS
ei	490,5
jah	2105,1
nii ja naa	716
(blank)	68
Grand Total	3379,6

5. Leidke *Exceli* poolt vaikimisi arvutatud kehamasside **summa asemel tudengite arv** erinevates gruppides.



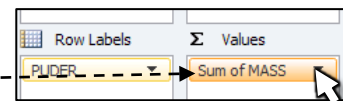
Selle osa võite vahele jätta

☞ **Alternatiivne variant** funktsiooni muutmiseks *Pivot Table*'s on klikkida *PivotTable Tools*-saki *Options*-alamsaki ikoonil *Summarize Values By* ning valida soovitud funktsioon sealt:



☞ Rakendatavat funktsiooni saab muuta ka klikkides muudetaval tunnusel

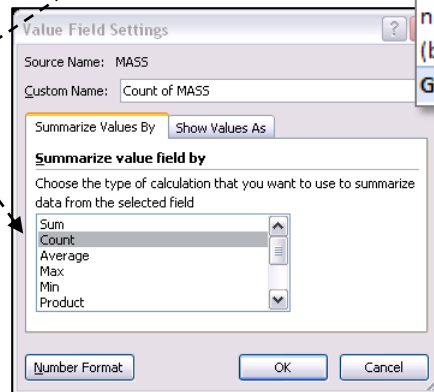
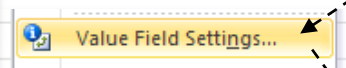
hiire vasakpoolse klahviga *Pivot Table*'i konstrueerimise aknas lahtris ' Σ Values'



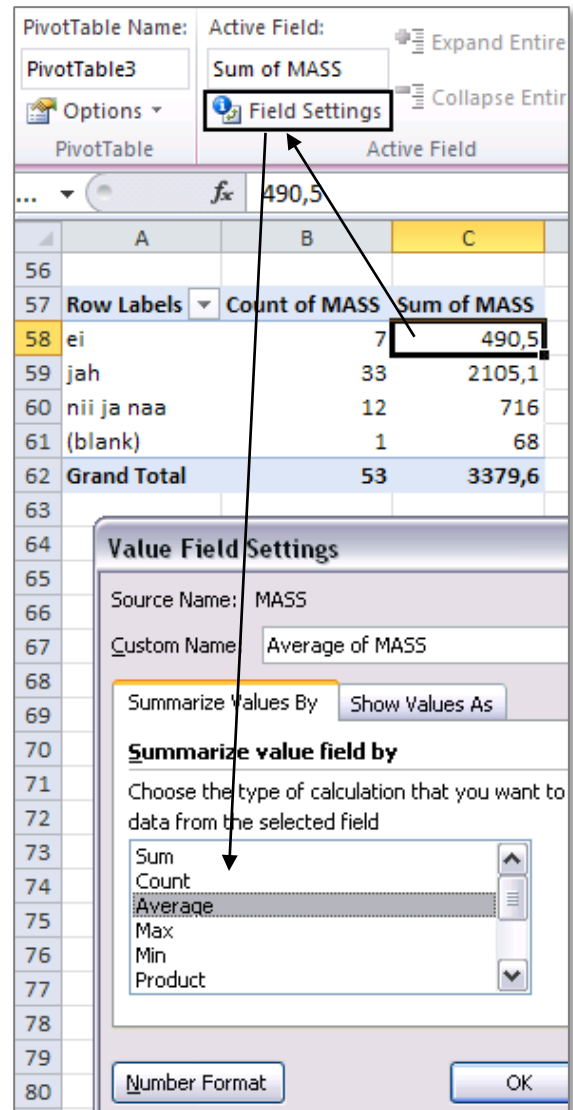
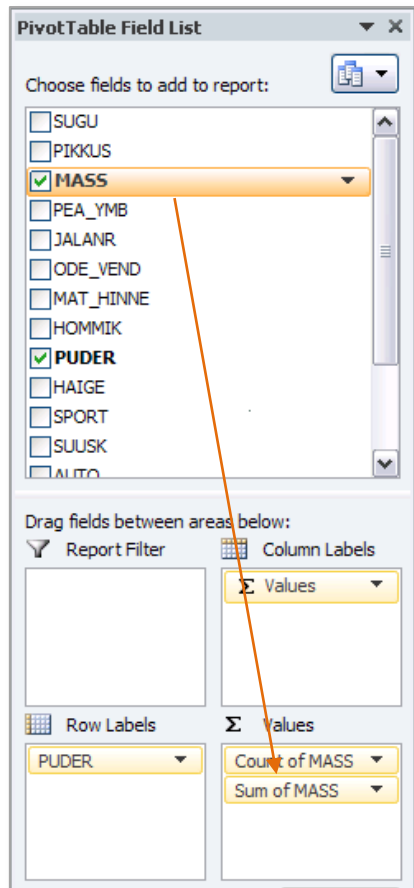
või hiire parempoolse klahviga *Pivot Table*'s

ning valides avanenud rippmenüüst käsu *Value Field Settings...* ja

sealt edasi vajaliku funktsiooni:



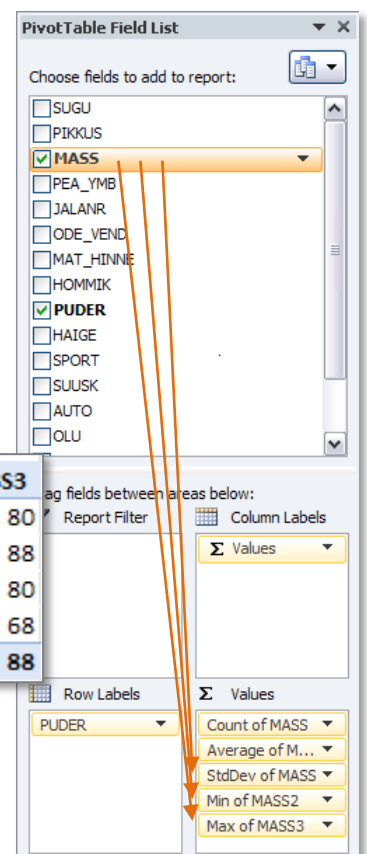
6. Arvutage lisaks ka keskmised kehamassid.



7. Arvutage ka kehamassi standardhälbed ning minimaalsed ja maksimaalsed väärtused.

Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	7	70,07142857	11,86180344	46,5	80
jah	33	63,79090909	10,43247225	47	88
nii ja naa	12	59,66666667	10,36895132	51	80
(blank)	1	68	#DIV/0!	68	68
Grand Total	53	63,76603774	10,74848725	46,5	88



Rida tähisega (**blank**) märgib seda tudengit, kelle pudrusöömise kohta info puudub (ta kas ei tea, mis asi puder on, või jäi tal antud küsimusele lihtsalt vastamata).

Sageli on mõttekas jätta taolised *Exceli* poolt automaatselt tekitatavad puuduvatele väärtustele vastavad grupid vaatluse alt välja:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	7	70,07142857	11,86180344	46,5	80
jah	33	63,79090909	10,43247225	47	88
nii ja naa	12	59,66666667	10,36895132	51	80
(blank)	1	68	#DIV/0!	68	68
Grand Total	53	63,76603774	10,74848725	46,5	88

8. Ümardage keskmised ja standardhälbed ühe kohani peale koma (näiteks vastava käsu abil *Home*-sakilt).

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	7	70,07142857	11,86180344	46,5	80
jah	33	63,79090909	10,43247225	47	88
nii ja naa	12	59,66666667	10,36895132	51	80
Grand Total	52	63,68461538	10,83683596	46,5	88

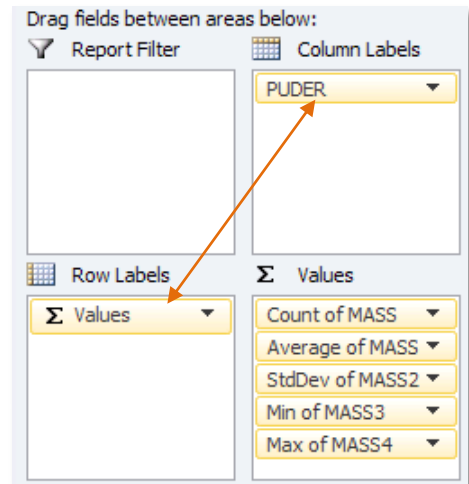
Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	7	70,1	11,9	46,5	80
jah	33	63,8	10,4	47	88
nii ja naa	12	59,7	10,4	51	80
Grand Total	52	63,7	10,8	46,5	88

9. Esitage viimati konstrueeritud tabel kujul, kus igale arvutatud suurusele vastab üks rida ja igale pudrusöömise väärtusele üks veerg:

Values	ei	jah	nii ja naa	Grand Total
Count of MASS	7	33	12	52
Average of MASS	70,1	63,8	59,7	63,7
StdDev of MASS	11,9	10,4	10,4	10,8
Min of MASS2	46,5	47	51	46,5
Max of MASS3	80	88	80	88

Spikker:

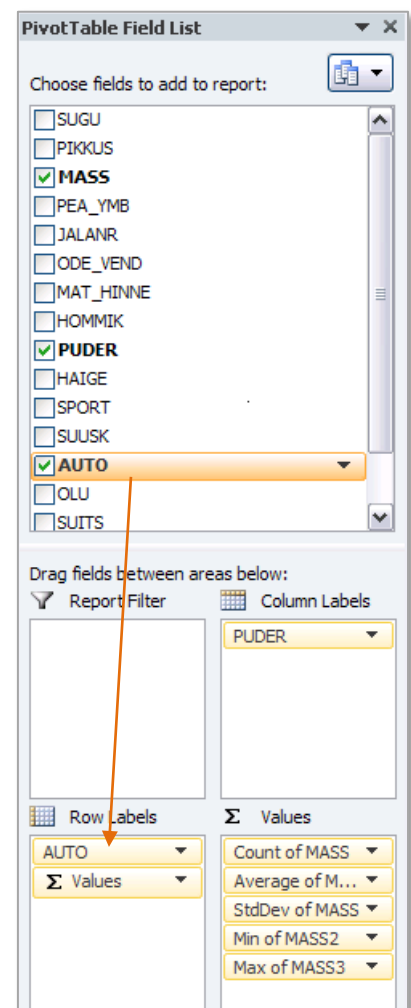


10. Ja nüüd jagage saadud tabel ridadeks auto omamist näitava tunnuse AUTO järgi.

Oodatav tulemus:

Row Labels	ei	jah	nii ja naa	Grand Total
ei				
Count of MASS	3	11	8	22
Average of MASS	61,5	59,5	61,6	60,6
StdDev of MASS	14,3	7,3	11,8	9,6
Min of MASS2	46,5	48	52	46,5
Max of MASS3	75	70	80	80
jah				
Count of MASS	4	22	4	30
Average of MASS	76,5	65,9	55,8	66,0
StdDev of MASS	4,0	11,2	6,4	11,3
Min of MASS2	73	47	51	47
Max of MASS3	80	88	65	88
Total Count of MASS	7	33	12	52
Total Average of MASS	70,1	63,8	59,7	63,7
Total StdDev of MASS	11,9	10,4	10,4	10,8
Total Min of MASS2	46,5	47	51	46,5
Total Max of MASS3	80	88	80	88

Spikker:



11. Proovige *Pivot Table* erinevaid esitusviise (mõni esitus võimaldab tabeli struktuurist paremini aru saada):

The screenshot shows the 'Report Layout' dropdown menu in Excel's PivotTable Tools ribbon. The menu is open, showing several options. The 'Show in Outline Form' option is highlighted in yellow. The background shows a PivotTable with the following data:

	E	F	G	H
56				
57				
58	Row Labels			Grand Total
59	ei			
60	Count of MASS	8	22	
61	Average of MASS	6	60,6	
62	StdDev of MASS	8	9,6	
63	Min of MASS2	2	46,5	
64	Max of MASS3	0	80	
65	jah			
66	Count of MASS	4	22	4
67	Average of MASS	76,5	65,9	55,8
68	StdDev of MASS	4,0	11,2	6,4
69	Min of MASS2	73	47	51
70	Max of MASS3	80	88	65
71	Total Count of MASS	7	33	12
72	Total Average of MASS	70,1	63,8	59,7
73	Total StdDev of MASS	11,9	10,4	10,8
74	Total Min of MASS2	46,5	47	51
75	Total Max of MASS3	80	88	80

12. Oskate te iga viimases tabelis sisalduva arvu kohta öelda, mida see näitab?

Proovige sõnastada paar järeldust keskmise kehamassi sarnasusest või erinevusest

- autot omavatel ja mitte omavatel ning
- putru söövatel ja mitesöövatel tudengitel.