

Biomeetria praks 2

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excel*'is 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 Excel*'is 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 *Excel*'i 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 *Excel*'ile 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 $n = 56$ (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaübermõõtude väärtused võiks jagada 7 või 8 klassi (sest $\sqrt{56} \approx 7,5$; 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 *Excel*'i 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	N	O	P	SUN
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANF	ODE_VEN	MAT_HI	HOMMIK	PUDER	LEMMIK	HAIGE	SPORT	SUUSK	AUTO	OLU	SUN
2	Eesti	N	168	62	53	40	1	5	helbed või müslid	jah	jah	jah	jah	jah	ei		0 ei
3	Eesti	N	178	75	56	40	4	5	helbed või müslid	jah	jah	ei	ei	ei	ei		0 jah
4	Eesti	N	170,5	64	54	38	2	4	võileib	nii ja naa	jah	ei	jah	ei	ei		0 ei
5	Eesti	N	158	56	50	39	2	3	ei söö tavaliselt h	jah	jah	ei	jah	jah	ei		0 ei
6	Eesti	N	163	60	54	38	1	4	puder	jah	jah	jah	jah	jah	ei		0 ei
7	Eesti	N	177	70	53	39	3	4	helbed või müslid	jah	jah	ei	jah	jah	jah		0 ei
8	Eesti	N	170	51	52	40	0	4	muu	nii ja naa	jah	ei	jah	jah	jah		0 ei
9	Eesti	N	170	56	38	39	2	4	helbed või müslid	jah	jah	ei	jah	jah	jah		0 ei
10	Soome	M	183	90	60	43	2	4	võileib	jah	ei	ei	jah	jah	jah		4 jah
11	Soome	N	167	60	54	40	1	3	võileib	nii ja naa	jah	ei	jah	ei	jah		0 ei
12	Soome	N	157	46	54	37	2	5	võileib	nii ja naa	ei	jah	jah	ei	ei		0 ei
13	Soome	N	175	65	56	40	2	3	puder	jah	ei	ei	jah	jah	ei		0 ei
14	Soome	N	157	70	56	36	2	3	puder	jah	ei	jah	jah	jah	ei		1 enar
15	Soome	N	173	62	56	39											ei
16	Soome	N	161	65	58	39											jah
17	Eesti	N	169	68	55,5	39											ei
18	Eesti	N	178	56	56	40											jah
19	Eesti	N	164	62	55	37,5											ei
20	Eesti	N	159	55	54	37											jah
21	Eesti	N	164	55	54,5	37											ei
22	Eesti	N	174	74	58	40											jah
23	Eesti	N	171	67	66	38											ei
24	Soome	N	160	56	57	38											ei
25	Soome	N	170	76	56	39											jah
26	Soome	M	187	75	55	44											ei
27	Soome	N	165	55	56	37											ei
28	Eesti	N	168	55	57	39											ei
29	Eesti	N	159	48	52	34											ei
30	Soome	N	166	58	57	38											jah
31	Soome	N	169	59	58	38											ei
32	Soome	N	160	57	58	37											jah
33	Soome	N	178	70	53	41											ei
34	Soome	N	169	68	58	40											ei
35	Eesti	N	174	62	58	40											ei
36	Eesti	N	165	59	55	37											jah
37	Eesti	N	177	76	55	40											ei
38	Soome	N	170	80	59	39											jah
39	Eesti	N	173	75	55	40											ei
40	Eesti	N	175,5	66	55	39											ei
41	Eesti	N	172	66	56	40											ei
42	Eesti	N	165	60	57	39											ei
43	Eesti	N	170	60	58	38											ei
44	Eesti	N	168	55	58	39											ei
45	Soome	N	158	75	58	38											ei
46	Eesti	N	170	54	53	39											ei
47	Soome	N	181	72	57	42											jah
48	Soome	N	167	70	51,5	40											ei
49	Soome	N	166	49	55	38											ei
50	Soome	N	174	51	55	38											ei
51	Soome	N	160	60	57	38											jah
52	Soome	N	170	65	59	39											jah
53	Soome	N	170	66	58	38											ei
54	Eesti	M	171	75	56	39											ei
55	Eesti	M	177	69	55	43											ei
56	Eesti	N	168	65	55	39											jah
57	Eesti	N	172	62	53	39											ei
58																	
59																	

- Maksimaalne peaübermõõt tuleb 66 ja minimaalne 38 cm. Kas need arvud on reaalsed? Kui übermõõtu on omale keeruline ette kujutada, siis diameetriga (läbimõõduga) on ehk lihtsam. Siit ka küsimus – kui suur on pea läbimõõt, kui übermõõt on 38 cm, ja kas see on reaalne?

Spikker. Mäletatavasti avaldub ringi übermõõt valemiga $P = 2\pi r$, millest diameeter $d = 2r = P/\pi$ (Excelit arvutusmasina kasutades võite sisestada valemi kujul '=38/PI()').

Tulemuseks peaks tulema ~12,1 cm. See on umbes sama palju, kui tavalisel CD-plaadil. Seega tundub peaübermõõt 38 cm pisut liiga väike.

Ilmselt on tegu mõõtmis- või andmete ülesmärkimisveaga.

Et traditsioonilised statistikameetodid ei ole mõeldud ebareaalsete või väga erandlike andmete analüüsimiseks, jäetakse sellised väärtused tavaliselt analüüsist välja. Antud juhul on kõige mõistlikum peaübermõõdu väärtus 38 cm andmetabelist lihtsalt ära kustutada.

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

51	52	4L	51	52	4L
56	38	3E	56		3E
90	60	4E	90	60	4E

- Minimaalne peaümberrõõd korregeeritud andmetest on 50 ja maksimaalne 66 cm.
- Klasside intervallide ligikaudseks leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovivate klasside arvuga ehk $(66-50)/8 = 2$ cm.

Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli vajadusel veidi suurendada või vähendada (antud ülesande puhul ei ole see vajalik, sest 2 cm on juba täisarvuline väärtus).

- Klasside piiride paika panekul alustatakse esimesest klassist, mis määratakse nii, et see sisaldaks minimaalset väärtust, näiteks võiks esimene klass olla 48-50 cm.
- Kokkuvõttes võime tudengite peaümberrõõdudest moodustada üheksa 2-sentimeetrist klassi:

48-50; 50-52; 52-54; 54-56; 56-58; 58-60; 60-62; 62-64; 64-66.

Et *Excel* paigutab klassi piiriga võrdsed väärtused alumisse klassi, on korrektsem esitada klassid poollõikudena (sest kuidas muidu aru saada, kumba klassi näiteks 50 cm kuulub?):

(48,50], (50,52], (52,54], (54,56], (56,58], (58,60], (60,62], (62,64] ja (64,66].

NB. Soovi korral võinuks moodustada ka hoopis **seitse 3-sentimeetrist** klassi kujul

(47,50], (50,53], (53,56], (56,59], (59,62], (62,65], (65,68];

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

(48,51], (51,54], (54,57], (57,60], (60,63], (63,66].

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

R	S	T	U
	KINO		pea_ymb klassid
ui aasta tagasi	viimase kuu jooksul		50
kuu jooksul	viimase kuu jooksul		52
aasta jooksul	viimase aasta jooksul		54
ui aasta tagasi	viimase 10 päeva jooksul		56
aasta jooksul	viimase aasta jooksul		58
aasta jooksul	viimase kuu jooksul		60
aasta jooksul	rohkem kui aasta tagasi		62
ui aasta tagasi	viimase aasta jooksul		64
kuu jooksul	viimase aasta jooksul		

- Märkusi:
 - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;
 - *Excel* tõlgendab etteantud väärtusi järgmiselt:
 - '50' tähendab ' ≤ 50 ';
 - '52' tähendab ' $50 < x \leq 52$ ' 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 ' 66 '= $(64,66]$ 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 64-st suuremad väärtused.

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

The screenshot shows the Excel interface with the 'Data Analysis' dialog box open. The 'Histogram' option is highlighted in the list of analysis tools. A red dashed line traces the path from the 'Data' tab in the ribbon to the 'Data Analysis' button, and then to the 'Histogram' option in the dialog box. The background spreadsheet contains data for various variables like 'PIKKUS', 'MASS', 'PEA_P', etc.

The screenshot shows the 'Histogram' dialog box with the following settings:

- Input Range:** \$E\$1:\$E\$57
- Bin Range:** \$U\$1:\$U\$9
- Labels**
- Output Range:** \$X\$1
- Cumulative Percentage**

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

Protseduuri *Histogram* oodatav tulemus:

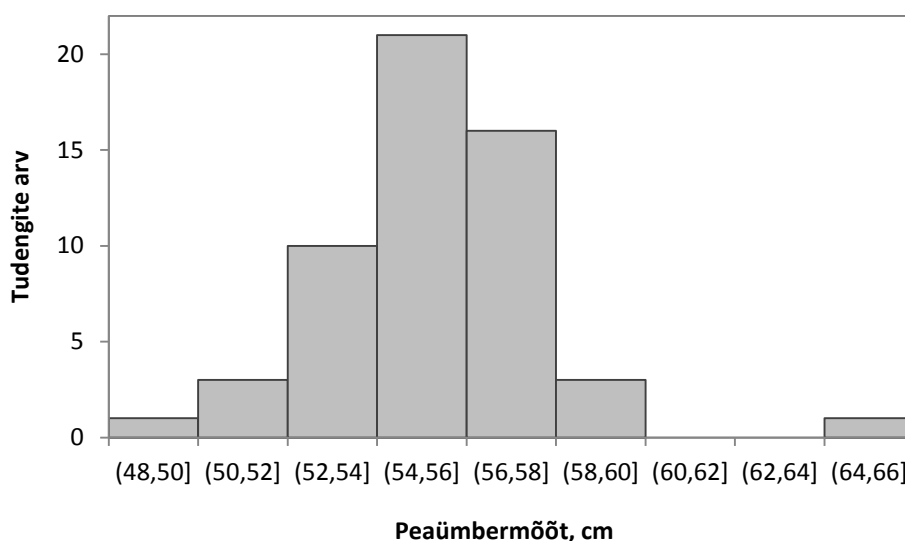
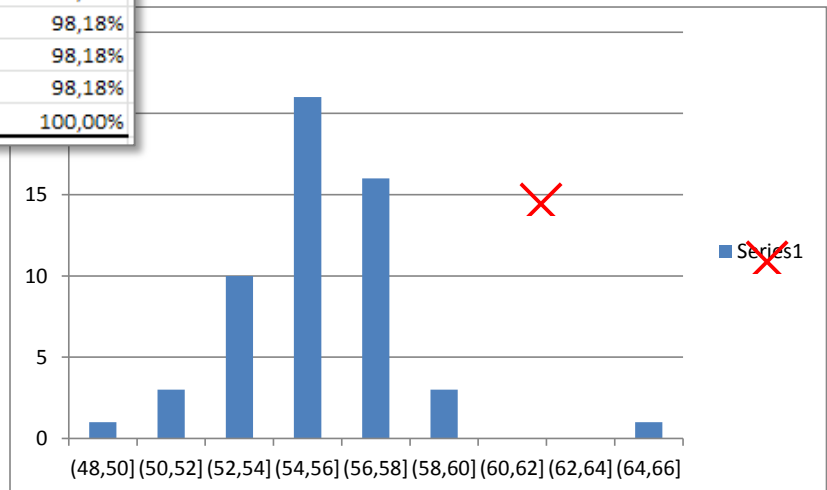
(kumulatiivne sagedus ehk jaotus näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

X	Y	Z
pea_ymb klassid	Frequency	Cumulative %
50	1	1,82%
52	3	7,27%
54	10	25,45%
56	21	63,64%
58	16	92,73%
60	3	98,18%
62	0	98,18%
64	0	98,18%
More	1	100,00%

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

Peaüumber	pea_ymb klassid	Frequency	Cumulative %
(48,50]	50	1	1,82%
(50,52]	52	3	7,27%
(52,54]	54	10	25,45%
(54,56]	56	21	63,64%
(56,58]	58	16	92,73%
(58,60]	60	3	98,18%
(60,62]	62	0	98,18%
(62,64]	64	0	98,18%
(64,66]	More	1	100,00%

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



- Kustutage legend ja ruudujooned
- Kaotage ära hall kast joonise ümbert ning lisage selle asemel hall kast ümber diagrammiala
- Y-telje ulatuseks 0 kuni 22 sammuga 5
- Pealkirjad x- ja y-teljele
- Tulpade vahe = 0
- Telgede ühikud ja nimed kirjasuuruses 10
- Tulpade 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.

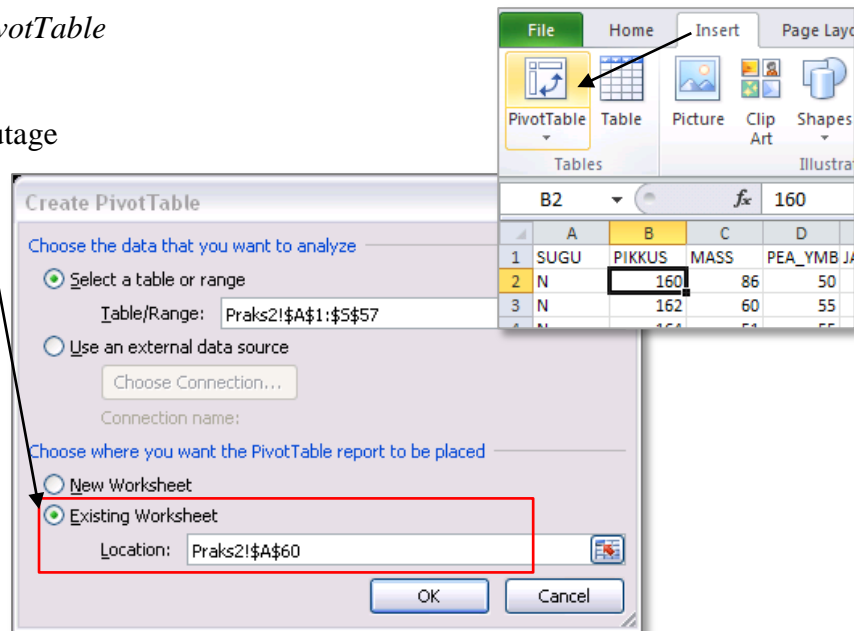
Ülesanne 2.

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

Tööjuhend

1. Paigutage kursor andmetabeli suvalisse lahtrisse ('Praks2'-lehel).
2. *Insert*-sakk → *PivotTable*

3. Loodav tabel paigutage 'Praks2'-lehele andmetabeli alla:

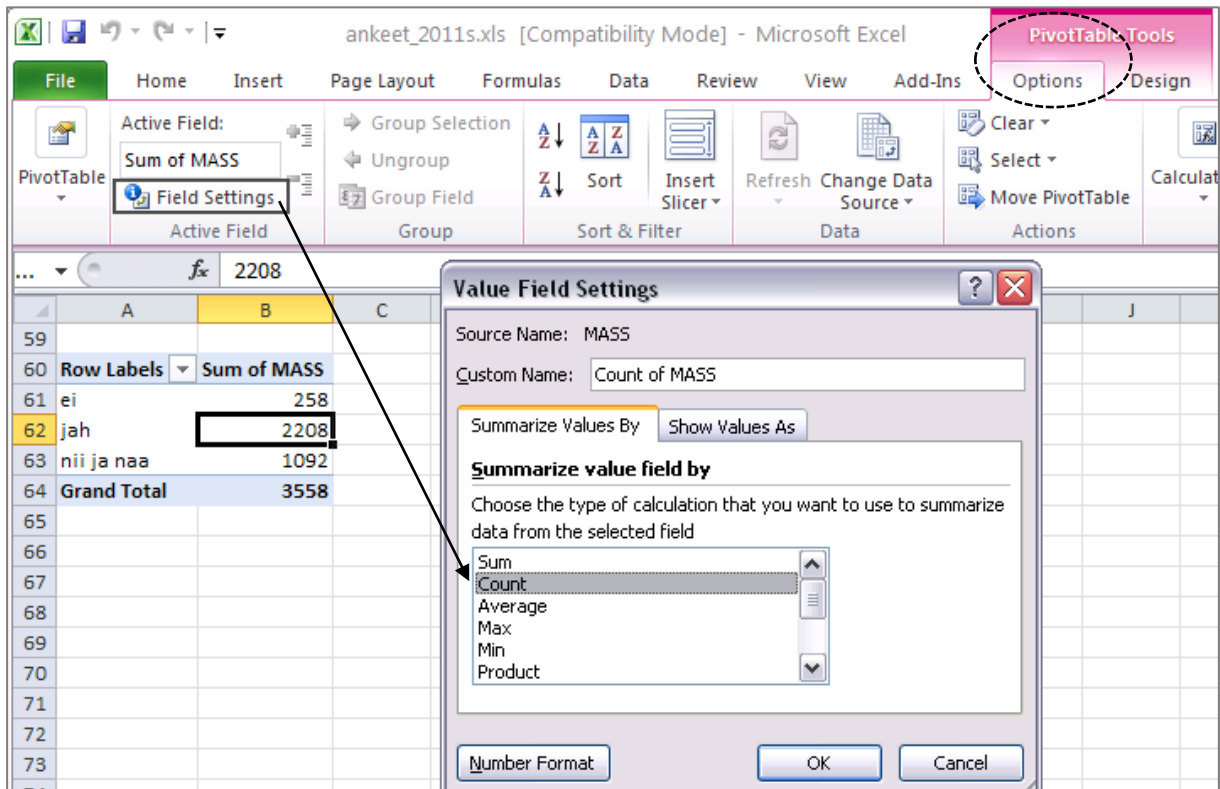


- 4.

Tulemuseks saadud tabel:

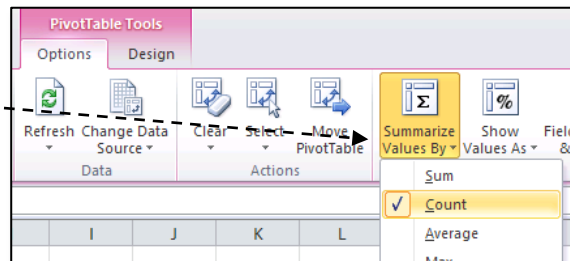
Row Labels	Sum of MASS
ei	258
jah	2208
nii ja naa	1092
Grand Total	3558

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



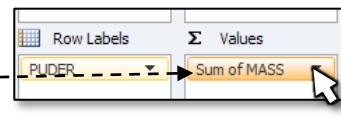
Selle osa võite vahele jätta

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



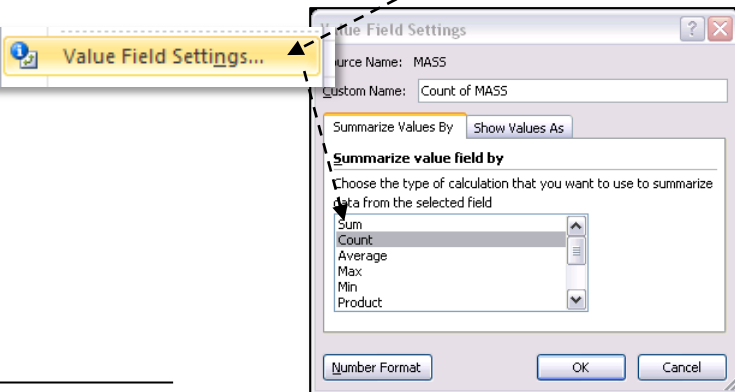
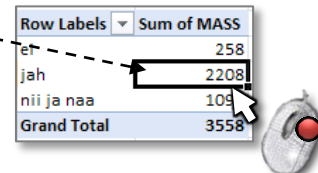
⌘) Rakendatavat funktsiooni saab muuta ka klikkides muudetaval tunnusel

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.

The screenshot shows the Excel PivotTable interface. On the left, the PivotTable Field List has 'MASS' and 'PUDER' checked. The PivotTable itself has 'Row Labels' as 'PUDER' and 'Values' as 'Sum of MASS'. The value for 'jah' is 2208. The 'Value Field Settings' dialog is open, with 'Average of MASS' selected under 'Summarize value field by'.

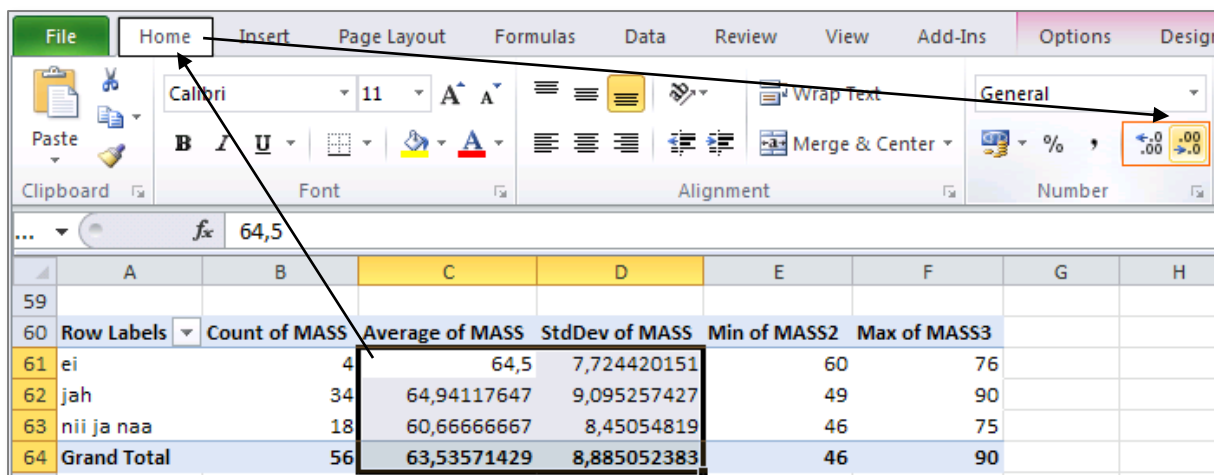
7. Arvutage ka kehamassi standardhälbed ning minimaalsed ja maksimaalsed väärtused.

Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	4	64,5	7,724420151	60	76
jah	34	64,94117647	9,095257427	49	90
nii ja naa	18	60,66666667	8,45054819	46	75
Grand Total	56	63,53571429	8,885052383	46	90

The screenshot shows the PivotTable Field List with 'MASS' and 'PUDER' selected. The 'Values' area contains five summary fields: 'Count of MASS', 'Average of MASS', 'StdDev of MASS', 'Min of MASS2', and 'Max of MASS3'.

8. Ümardage keskmised ja standardhälbed 1 kohani peale koma (näiteks vastava käsu abil *Home*-sakilt).



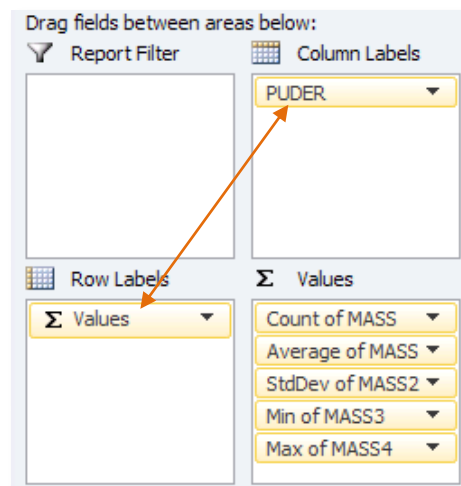
Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	4	64,5	7,7	60	76
jah	34	64,9	9,1	49	90
nii ja naa	18	60,7	8,5	46	75
Grand Total	56	63,5	8,9	46	90

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

Values	ei	jah	nii ja naa	Grand Total
Count of MASS	4	34	18	56
Average of MASS	64,5	64,9	60,7	63,5
StdDev of MASS	7,7	9,1	8,5	8,9
Min of MASS2	60	49	46	46
Max of MASS3	76	90	75	90

Spikker:

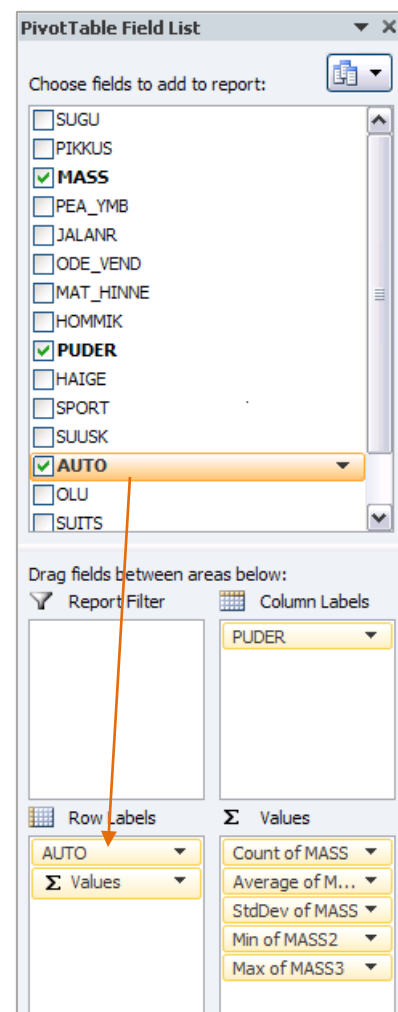


10. Ja nüüd jagage saadud tabel ridadeks ka veel auto omamist näitava tunnuse AUTO järgi.

Oodatav tulemus:

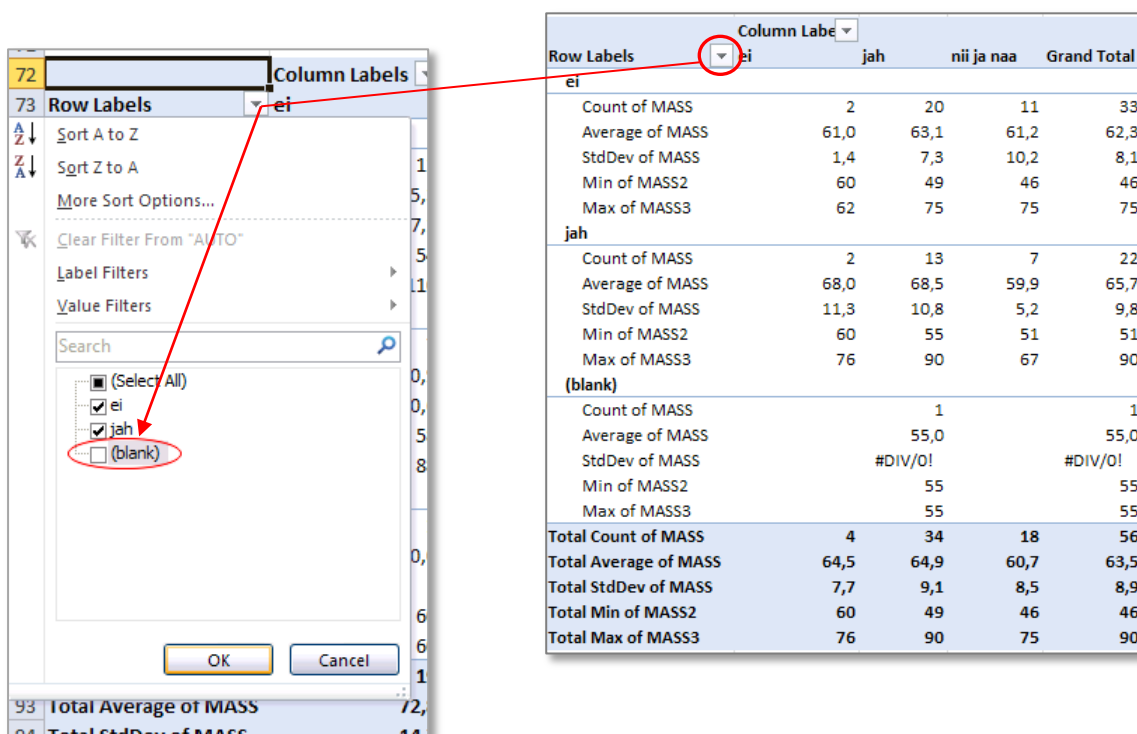
Row Labels	ei	jah	nii ja naa	Grand Total
ei				
Count of MASS	2	20	11	33
Average of MASS	61,0	63,1	61,2	62,3
StdDev of MASS	1,4	7,3	10,2	8,1
Min of MASS2	60	49	46	46
Max of MASS3	62	75	75	75
jah				
Count of MASS	2	13	7	22
Average of MASS	68,0	68,5	59,9	65,7
StdDev of MASS	11,3	10,8	5,2	9,8
Min of MASS2	60	55	51	51
Max of MASS3	76	90	67	90
(blank)				
Count of MASS		1		1
Average of MASS		55,0		55,0
StdDev of MASS		#DIV/0!		#DIV/0!
Min of MASS2		55		55
Max of MASS3		55		55
Total Count of MASS	4	34	18	56
Total Average of MASS	64,5	64,9	60,7	63,5
Total StdDev of MASS	7,7	9,1	8,5	8,9
Total Min of MASS2	60	49	46	46
Total Max of MASS3	76	90	75	90

Spikker:



Rida ja tabeli osa tähisega **(blank)** märgib nende tudengite gruppi, kelle auto kasutamise võimaluste kohta info puudub (nad kas ei tea, mis asi auto on, või jäi neil antud küsmusele lihtsalt interneti probleemide tõttu vastamata).

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



Tulemus:

Row Labels	Column Labels			Grand Total
ei	jah	nii	ja	naa
Count of MASS	2	20	11	33
Average of MASS	61,0	63,1	61,2	62,3
StdDev of MASS	1,4	7,3	10,2	8,1
Min of MASS2	60	49	46	46
Max of MASS3	62	75	75	75
jah				
Count of MASS	2	13	7	22
Average of MASS	68,0	68,5	59,9	65,7
StdDev of MASS	11,3	10,8	5,2	9,8
Min of MASS2	60	55	51	51
Max of MASS3	76	90	67	90
Total Count of MASS	4	33	18	55
Total Average of MASS	64,5	65,2	60,7	63,7
Total StdDev of MASS	7,7	9,1	8,5	8,9
Total Min of MASS2	60	49	46	46
Total Max of MASS3	76	90	75	90

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

The screenshot shows the Microsoft Excel interface with the PivotTable Tools ribbon active. The 'Report Layout' dropdown menu is open, displaying several options. The 'Show in Outline Form' option is highlighted in yellow. Below the menu, a PivotTable is visible, showing data for 'ei' and 'jah' categories. The PivotTable has columns for 'jah', 'nii ja naa', and 'Grand Total'. The data is summarized as follows:

Row Labels	jah	nii ja naa	Grand Total
Count of MASS	2	11	33
Average of MASS	61,0	61,2	62,3
StdDev of MASS	1,4	10,2	8,1
Min of MASS2	60	46	46
Max of MASS3	62	75	75
jah			
Count of MASS	2	7	22
Average of MASS	68,0	59,9	65,7
StdDev of MASS	11,3	5,2	9,8
Min of MASS2	60	51	51
Max of MASS3	76	67	90
Total Count of MASS	4	18	55
Total Average of MASS	64,5	60,7	63,7
Total StdDev of MASS	7,7	8,5	8,9
Total Min of MASS2	60	46	46
Total Max of MASS3	76	75	90

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
- kaerahelbeputru söövatel ja mitte söövatel tudengitel.