

## 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\_YMB' (peaümberrõõm) 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  $n = 68$  (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaümberrõõmide väärtused võiks jagada 8 klassi (sest  $\sqrt{68} \approx 8,2$ ; samas ei ole see absoluutne tõde, kui klasside piiridega paremini sobib, võib väärtused jagada ka 7 või hoopis 9 või 10 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.

The screenshot shows an Excel spreadsheet with a 'Customize Status Bar' dialog box open. The dialog box has several options checked, including 'Cell Mode', 'Signatures', 'Information Management Policy', 'Permissions', 'Fixed Decimal', 'End Mode', 'Page Number', 'Average', 'Count', 'Numerical Count', 'Minimum', 'Maximum', 'Sum', 'Upload Status', 'View Shortcuts', 'Zoom', and 'Zoom Slider'. The 'Average' option is highlighted with a red box, showing a value of 55,07352941. The 'Minimum' option is also highlighted with a yellow box, showing a value of 49. The status bar at the bottom of the spreadsheet shows 'Average: 55,07352941 Count: 68 Numerical Count: 68 Min: 49 Max: 63'. A mouse cursor is pointing at the 'Average' option in the dialog box.

- Maksimaalne peaümberrõõm tuleb 63 ja minimaalne 49 cm.
- Klasside intervallide leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovitatavate klasside arvuga ehk  $(63-49)/8 \approx 1,75$  cm.  
Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli veidi suurendada või vähendada, antud ülesande puhul sobiks intervalli pikkuseks ilmselt 2 cm.
- 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 peaümberrõõtudest moodustada seitse 2-sentimeetrist klassi:

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

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?):

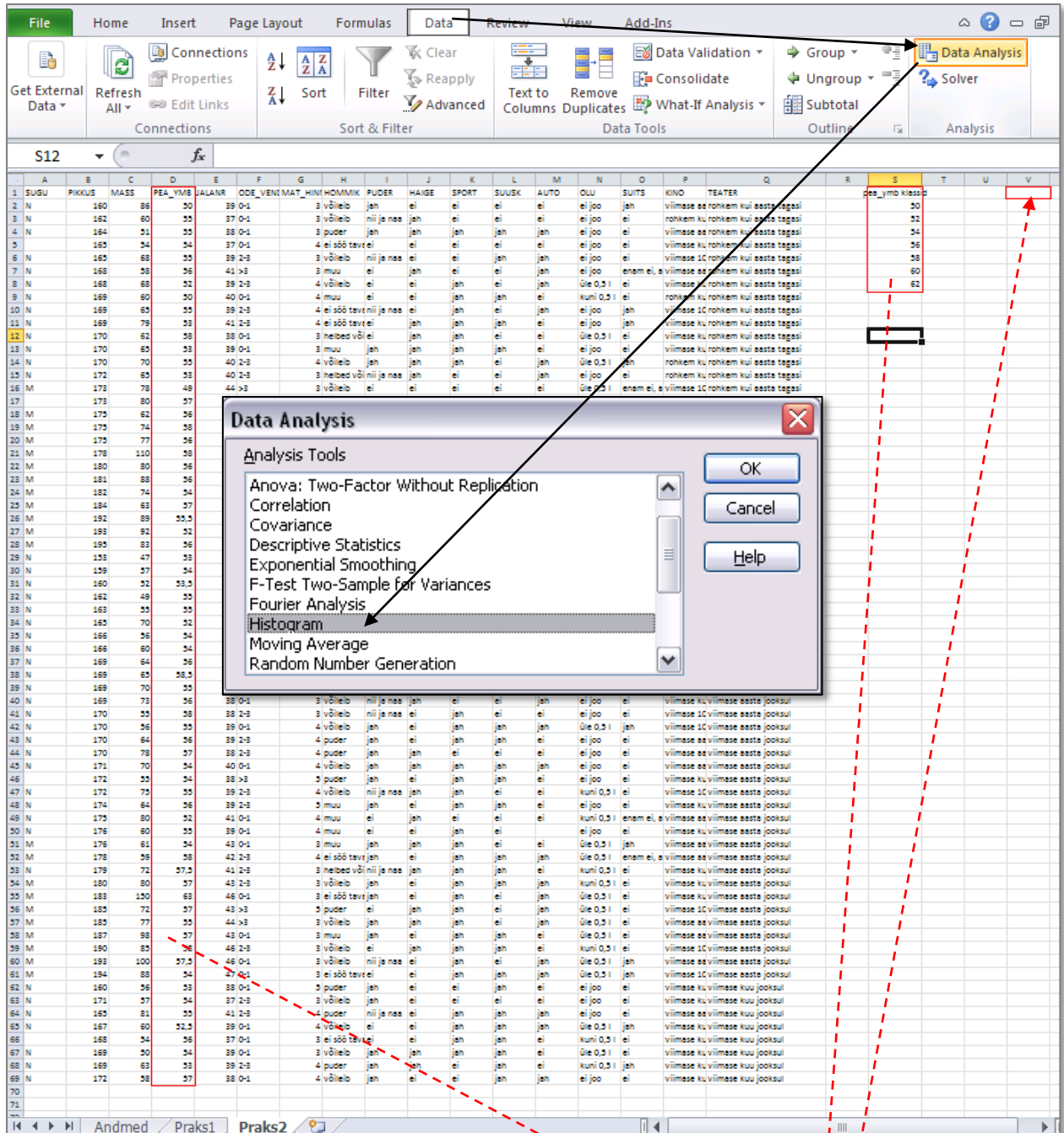
(48,50], (50,52], (52,54], (54,56], (56,58], (58,60], (60,62] ja (62,64].

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

J	SUITS	KINO	TEATER	pea_ymb klassid
100	jah	viimase	rohkem kui aasta tagasi	50
100	ei	rohkem	rohkem kui aasta tagasi	52
100	ei	viimase	rohkem kui aasta tagasi	54
100	ei	viimase	rohkem kui aasta tagasi	56
100	ei	viimase	rohkem kui aasta tagasi	58
100	enam ei	viimase	rohkem kui aasta tagasi	60
10,5	ei	viimase	rohkem kui aasta tagasi	62
10,5	ei	rohkem	rohkem kui aasta tagasi	

- Märkusi:
  - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;
  - *Excel* tõlgendab etteantud väärtusi järgmiselt:
    - '50' tähendab ' $\leq 50$ ';
    - '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 '64'=(62,64] 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 62-st suuremad väärtused.

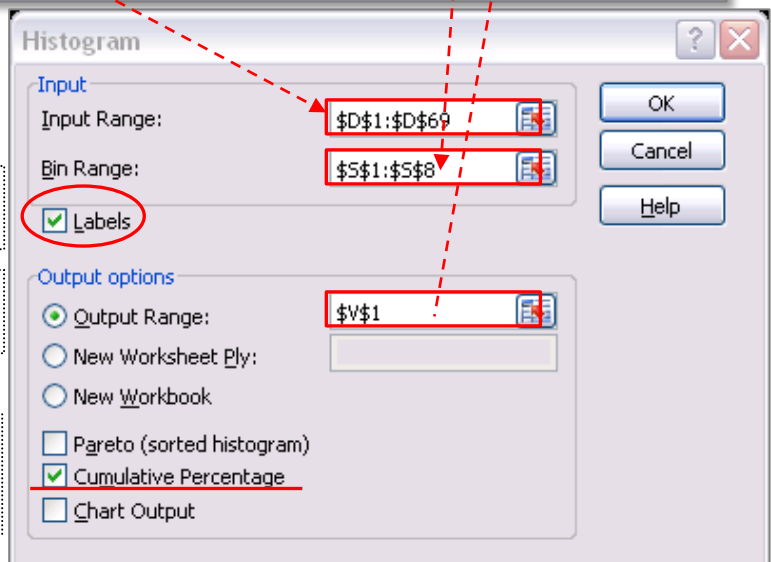
2. Sagedustabel protseduuri *Histogram* abil: *Data*-sakk → *Data Analysis...* → *Histogram*



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:

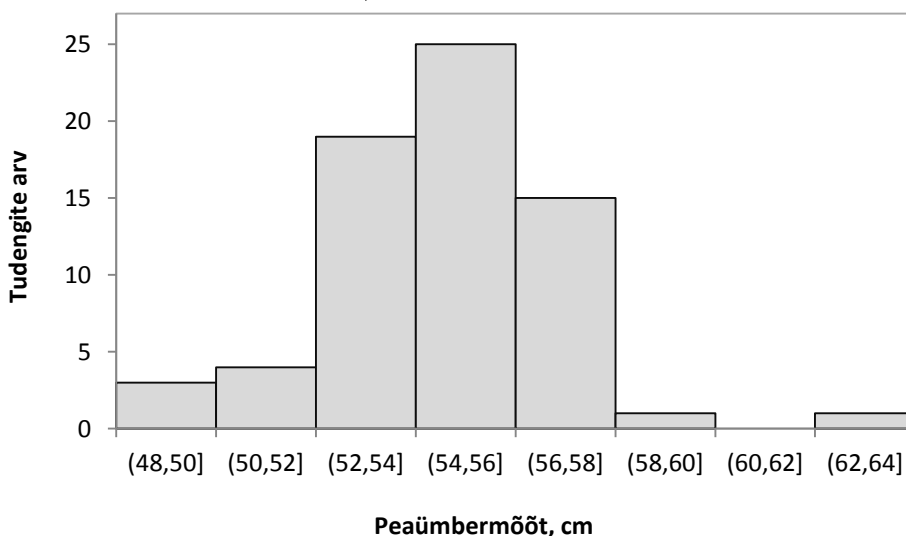
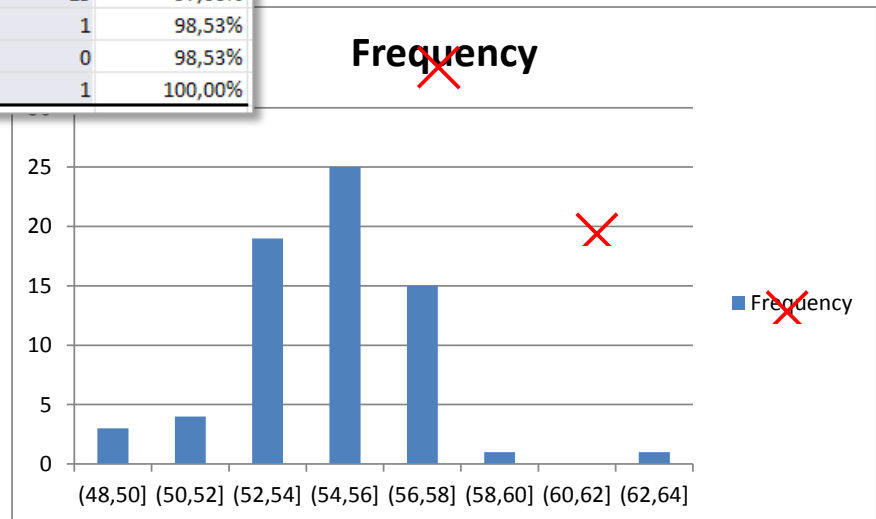
(kumulatiivne sagedus ehk jaotus näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

V	W	X
pea_ymb klassid	Frequency	Cumulative %
50	3	4,41%
52	4	10,29%
54	19	38,24%
56	25	75,00%
58	15	97,06%
60	1	98,53%
62	0	98,53%
More	1	100,00%

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

Peaümberrõõt	pea_ymb klassid	Frequency	Cumulative %
(48,50]	50	3	4,41%
(50,52]	52	4	10,29%
(52,54]	54	19	38,24%
(54,56]	56	25	75,00%
(56,58]	58	15	97,06%
(58,60]	60	1	98,53%
(60,62]	62	0	98,53%
(62,64]	More	1	100,00%

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



- Kustutage pealkiri, legend ja ruudujooned
- Kaotage ära hall kast joonise ümber ning lisage selle asemel hall kast ümber diagrammiala
- Pealkirjad x- ja y-teljele
- Tulpade vahe = 0
- Telgede ühikud ja nimed kirjasuuruses 10
- Tulbad 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.

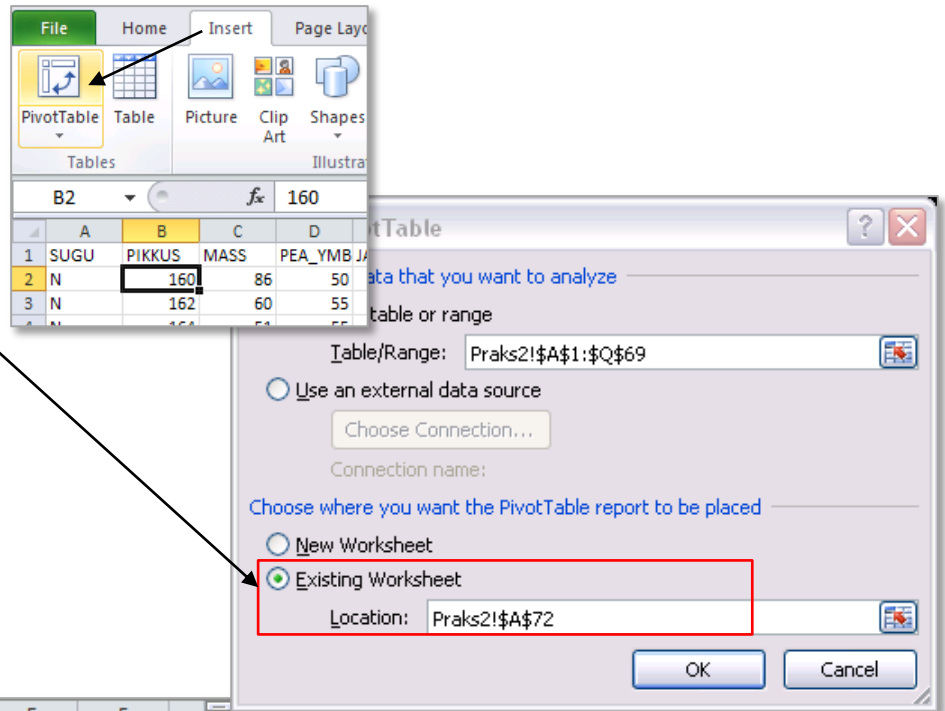
## Ülesanne 2.

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:

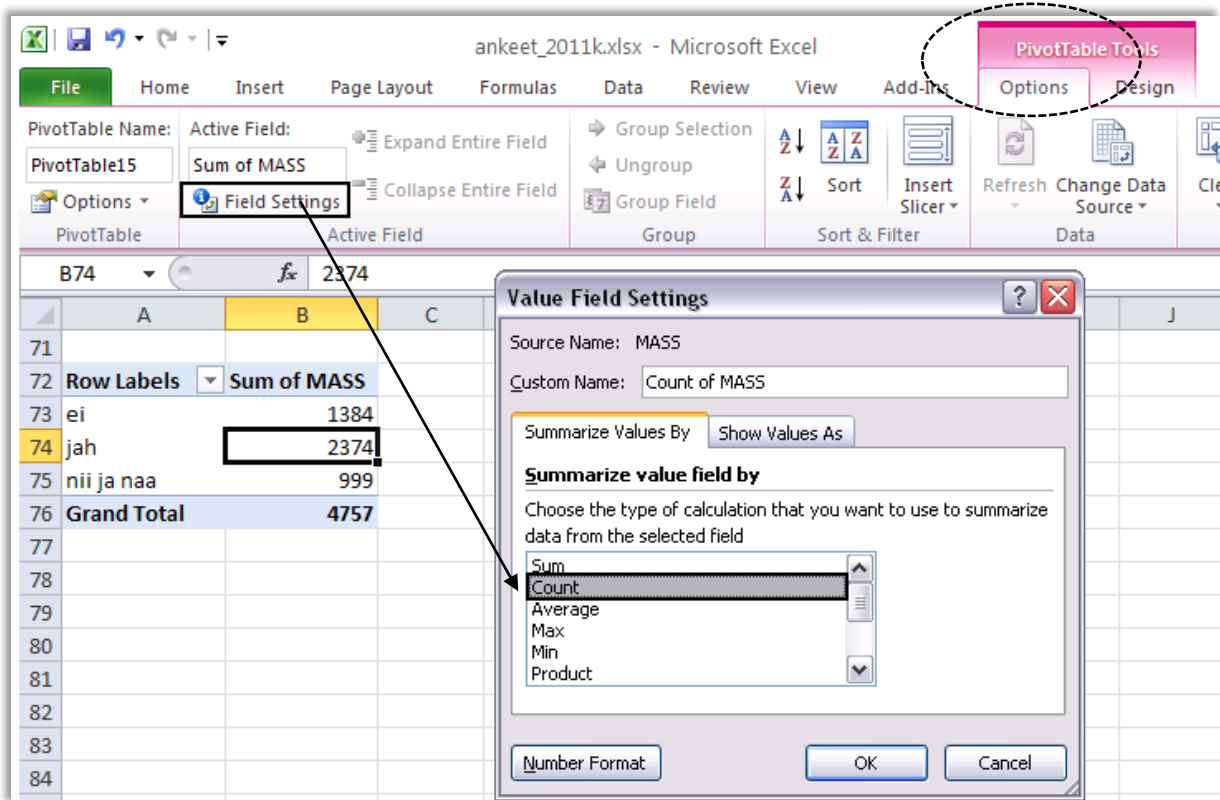


- 4.

Tulemuseks saadud tabel:

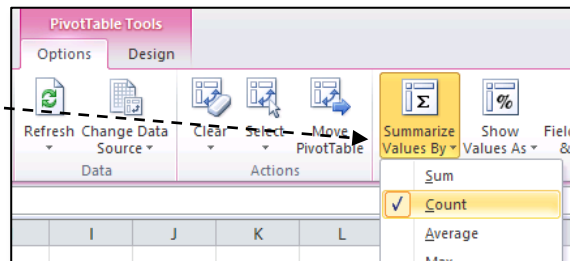
Row Labels	Sum of MASS
ei	1384
jah	2374
nii ja naa	999
<b>Grand Total</b>	<b>4757</b>

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



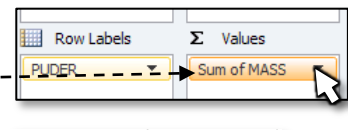
Selle osa võite vahele jätta

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



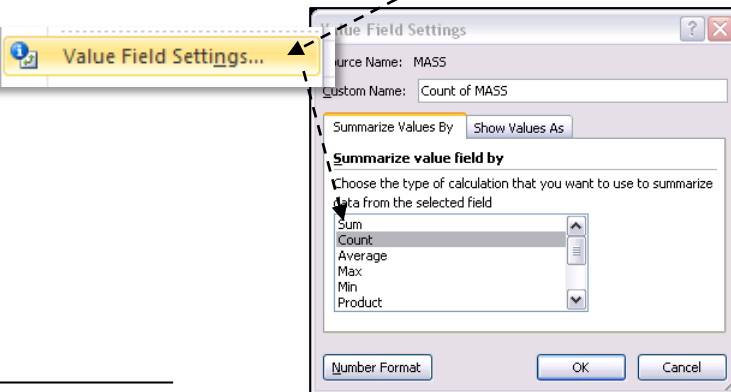
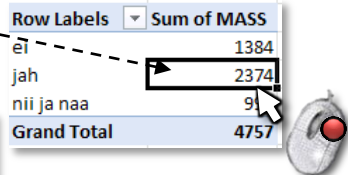
☞) Rakendatavat funktsiooni saab muuta ka klikkides muudetaval tunnusel

**hiire vasakpoolse klahviga** *Pivot Table*'i konstrueerimise aknas lahtris ' $\Sigma$  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' pane has 'MASS' and 'PUDER' checked. The main area shows a PivotTable with 'Row Labels' as 'ei', 'jah', 'nii ja naa', and 'Grand Total'. The 'Sum of MASS' column shows values 1384, 2374, 999, and 4757 respectively. The 'Value Field Settings' dialog is open, with 'Average of MASS' selected in the 'Summarize value field by' list.

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

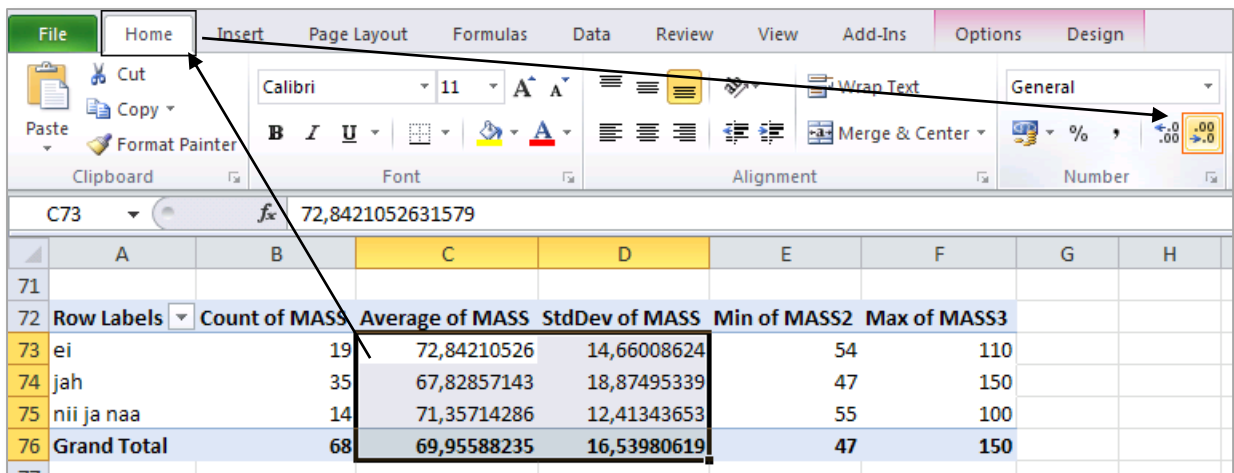
Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	19	72,84210526	14,66008624	54	110
jah	35	67,82857143	18,87495339	47	150
nii ja naa	14	71,35714286	12,41343653	55	100
<b>Grand Total</b>	<b>68</b>	<b>69,95588235</b>	<b>16,53980619</b>	<b>47</b>	<b>150</b>

The screenshot shows the 'PivotTable Field List' pane with 'MASS' and 'PUDER' selected. The 'Values' area contains five 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).



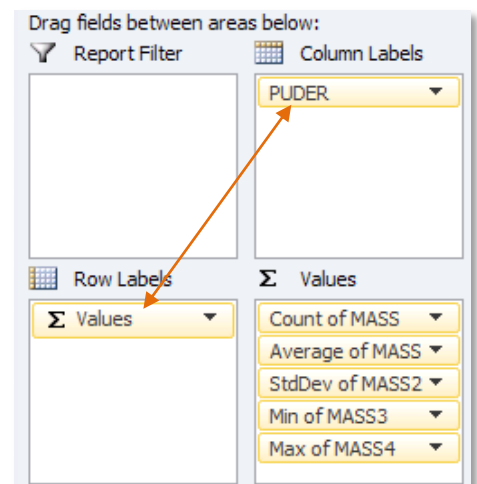
Tulemus:

Row Labels	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	19	72,8	14,7	54	110
jah	35	67,8	18,9	47	150
nii ja naa	14	71,4	12,4	55	100
<b>Grand Total</b>	<b>68</b>	<b>70,0</b>	<b>16,5</b>	<b>47</b>	<b>150</b>

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	19	35	14	68
Average of MASS	72,8	67,8	71,4	70,0
StdDev of MASS	14,7	18,9	12,4	16,5
Min of MASS2	54	47	55	47
Max of MASS3	110	150	100	150

Spikker:



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

Oodatav tulemus:

Column Labels	ei	jah	nii ja naa	Grand Total
<b>ei</b>				
Count of MASS	11	16	7	34
Average of MASS	75,3	66,1	70,0	69,9
StdDev of MASS	17,1	13,1	11,4	14,4
Min of MASS2	54	50	55	50
Max of MASS3	110	98	88	110
<b>jah</b>				
Count of MASS	7	17	7	31
Average of MASS	70,9	70,6	72,7	71,2
StdDev of MASS	10,6	23,9	14,2	19,2
Min of MASS2	58	47	57	47
Max of MASS3	88	150	100	150
<b>(blank)</b>				
Count of MASS	1	2		3
Average of MASS	60,0	57,5		58,3
StdDev of MASS	#DIV/0!	3,5		2,9
Min of MASS2	60	55		55
Max of MASS3	60	60		60
<b>Total Count of MASS</b>	<b>19</b>	<b>35</b>	<b>14</b>	<b>68</b>
<b>Total Average of MASS</b>	<b>72,8</b>	<b>67,8</b>	<b>71,4</b>	<b>70,0</b>
<b>Total StdDev of MASS</b>	<b>14,7</b>	<b>18,9</b>	<b>12,4</b>	<b>16,5</b>
<b>Total Min of MASS2</b>	<b>54</b>	<b>47</b>	<b>55</b>	<b>47</b>
<b>Total Max of MASS3</b>	<b>110</b>	<b>150</b>	<b>100</b>	<b>150</b>

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 puudevatele väärtustele vastavad grupid vaatluse alt välja jätta:

Tulemus:

Row Labels	Column Labels			Grand Total
ei	jah	nii ja naa		
ei				
Count of MASS	11	16	7	34
Average of MASS	75,3	66,1	70,0	69,9
StdDev of MASS	17,1	13,1	11,4	14,4
Min of MASS2	54	50	55	50
Max of MASS3	110	98	88	110
jah				
Count of MASS	7	17	7	31
Average of MASS	70,9	70,6	72,7	71,2
StdDev of MASS	10,6	23,9	14,2	19,2
Min of MASS2	58	47	57	47
Max of MASS3	88	150	100	150
<b>Total Count of MASS</b>	<b>18</b>	<b>33</b>	<b>14</b>	<b>65</b>
<b>Total Average of MASS</b>	<b>73,6</b>	<b>68,5</b>	<b>71,4</b>	<b>70,5</b>
<b>Total StdDev of MASS</b>	<b>14,7</b>	<b>19,3</b>	<b>12,4</b>	<b>16,7</b>
<b>Total Min of MASS2</b>	<b>54</b>	<b>47</b>	<b>55</b>	<b>47</b>
<b>Total Max of MASS3</b>	<b>110</b>	<b>150</b>	<b>100</b>	<b>150</b>

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 the following options:

- Show in Compact Form
- Show in Outline Form
- Show in Tabular Form
- Repeat All Item Labels
- Do Not Repeat Item Labels

The PivotTable data is visible in the background, showing the same data as the table above.

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.