

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 | RIK | SUGU | PIKKUS | MASS | PEA_P | JALANR | ODE_VEI | MAT_HIN | HOMMIK | PUDER | LEMMIK | HAIGE | SPORT |
| 2 | Eesti | N | 180 | 78 | 56 | 42 | 2 | 3 | muu | jah | ei | ei | jah |
| 3 | Eesti | N | 178 | 65 | 56 | 39 | 4 | 4 | ei söötav | jah | jah | ei | jah |
| 4 | Eesti | M | 177 | 70 | 57 | 42 | 3 | 3 | võileib | nii janaa | ei | jah | ei |
| 5 | Eesti | M | 187 | 75 | 45 | 45 | 3 | 3 | ei söötav | jah | jah | ei | jah |
| 6 | Eesti | M | 186 | 74 | 43 | 44 | 1 | 3 | helbed vi | jah | jah | ei | jah |
| 7 | Eesti | N | 165 | 62 | 42 | 37 | 3 | 3 | puder | jah | ei | jah | jah |
| 8 | Eesti | N | 170 | 67 | 55.5 | 40 | 4 | 3 | puder | jah | jah | jah | jah |
| 9 | Eesti | N | 177 | 59 | 42 | 39 | 1 | 4 | võileib | nii janaa | ei | ei | ei |
| 10 | Eesti | N | 166 | 47 | 55 | 38 | 2 | 4 | võileib | jah | jah | ei | jah |
| 11 | Eesti | N | 165 | 55 | 42 | 38 | 1 | 3 | võileib | jah | jah | ei | jah |
| 12 | Eesti | M | 180 | 68 | 51 | 44 | 0 | 3 | võileib | nii janaa | jah | ei | jah |
| 13 | Eesti | N | 161 | 49 | 56 | 37 | 2 | 3 | võileib | nii janaa | jah | jah | ei |
| 14 | Eesti | N | 168 | 54 | 50 | 40 | 0 | 5 | võileib | nii janaa | jah | ei | ei |
| 15 | Eesti | N | 167 | 67 | 56 | 40 | 2 | 3 | võileib | jah | jah | ei | jah |
| 16 | Eesti | N | 160 | 50 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 17 | Eesti | N | 164 | 53 | 59 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 18 | Eesti | M | 194 | 105 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 19 | Eesti | M | 178 | 65 | 53 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 20 | Eesti | M | 177 | 90 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 21 | Eesti | N | 171 | 57 | 50 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 22 | Eesti | M | 187 | 99 | 58 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 23 | Eesti | M | 189 | 81 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 24 | Eesti | M | 186 | 98 | 56 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 25 | Eesti | N | 171 | 65 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 26 | Eesti | M | 183 | 110 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 27 | Eesti | M | 193 | 100 | 58 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 28 | Eesti | N | 183 | 79 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 29 | Eesti | N | 177 | 75 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 30 | Eesti | N | 164 | 54 | 52 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 31 | Eesti | N | 170 | 60 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 32 | Eesti | N | 175 | 65 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 33 | Eesti | N | 168 | 69 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 34 | Eesti | M | 186 | 94 | 58 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 35 | Eesti | N | 160 | 55 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 36 | Eesti | N | 175 | 69 | 53 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 37 | Eesti | N | 151 | 53 | 52 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 38 | Eesti | M | 198 | 110 | 60 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 39 | Eesti | M | 174 | 120 | 56 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 40 | Eesti | N | 180 | 80 | 58 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 41 | Eesti | N | 171 | 72 | 30 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 42 | Eesti | N | 169 | 80 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 43 | Eesti | N | 160 | 64 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 44 | Eesti | N | 170 | 65 | 55 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 45 | Eesti | N | 176 | 58 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 46 | Eesti | N | 167 | 55 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 47 | Eesti | N | 160 | 59 | 56 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 48 | Eesti | N | 172 | 80 | 54 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 49 | Eesti | N | 170 | 60 | 30 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 50 | Eesti | N | 176 | 57 | 56 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 51 | Eesti | N | 168 | 62 | 56 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 52 | muu | M | 189 | 78 | 57 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 53 | Eesti | N | 173 | 68 | 49 | 38 | 0 | 4 | võileib | ei | ei | ei | ei |
| 54 | | | | | | | | | | | | | |

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 Not Recording
- Macro Recording
- Selection Mode
- Page Number
- Average 52.95192308
- Count 52
- Numerical Count 52
- Minimum 30
- Maximum 60
- Sum 2753.5
- Upload Status
- View Shortcuts
- Zoom 70%
- Zoom Slider

Ready | Average: 52.95192308 | Count: 52 | Numerical Count: 52 | Min: 30 | Max: 60 | Sum: 2753.5

- Maksimaalne peaübermõõt tuleb 60 ja minimaalne 30 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 diameeter, kui übermõõt on 30 cm, ja kas saadud tulemus 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 '=30/PI()'). Tulemuseks peaks tulema 9,55 cm. Võrdluseks – tavalise CD-plaadi läbimõõt on 12 cm. Seega tundub peaübermõõt 30 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 peaübermõõdu väärtus 30 cm andmetabelist lihtsalt ära kustutada (30-sentimeetriseid peaübermõõte on andmetabelis lausa kaks – ära kustutada tuleks need mõlemad).

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

| | | | | | |
|----|----|----|----|----|----|
| 80 | 58 | 42 | 80 | 58 | 42 |
| 72 | 30 | 41 | 72 | 30 | 41 |
| 80 | 57 | 40 | 80 | 57 | 40 |

- Minimaalne peäumbermõõt korrigeeritud andmetest on 42 ja maksimaalne 60 cm.
- Klasside intervallide ligikaudseks leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovitatavate klasside arvuga ehk $(60-42)/7 = 2,57$ cm.
Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli vajadusel veidi suurendada või vähendada. Antud juhul võiks näiteks valida klasside suuruseks 3 cm.
- Klasside piiride paika panekul alustatakse esimesest klassist, mis määratakse nii, et see sisaldaks minimaalset väärtust, näiteks võiks esimene klass olla 40-43 cm.
- Kokkuvõttes võime tudengite peäumbermõõtudest moodustada seitse 3-sentimeetrist klassi:

40-43; 43-46; 46-49; 49-52; 52-55; 55-58; 58-61.

Et *Excel* paigutab klassi piiriga võrdsed väärtused alumisse klassi, on korrektsem esitada klassid poollõikudena (sest kuidas muidu aru saada, kumba klassi näiteks 43 cm kuulub?):

(40,43], (43,46], (46,49], (49,52], (52,55], (55,58] ja (58,61].

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

(41,43], (43,45], (45,47], (47,49], (49,51], (51,53], (53,55], (55,57], (57,59], (59,61];

või siis hoopis **viis 4-sentimeetrist** klassi kujul

(40,44], (44,48], (48,52], (52,56], (56,60].

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

| S | T | U | V |
|---------------|---|---------|---------|
| | | pea_ymb | klassid |
| kuu jooksul | | | 43 |
| aasta jooksul | | | 46 |
| aasta jooksul | | | 49 |
| aasta jooksul | | | 52 |
| kuu jooksul | | | 55 |
| kuu jooksul | | | 58 |
| aasta jooksul | | | |

- Märkusi:
 - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;
 - *Excel* tõlgendab etteantud väärtusi järgmiselt:
 - '43' tähendab ' ≤ 43 ';
 - '46' tähendab ' $43 < x \leq 46$ ' 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 '61'=(58,61] 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 58-st suuremad väärtused.

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

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:

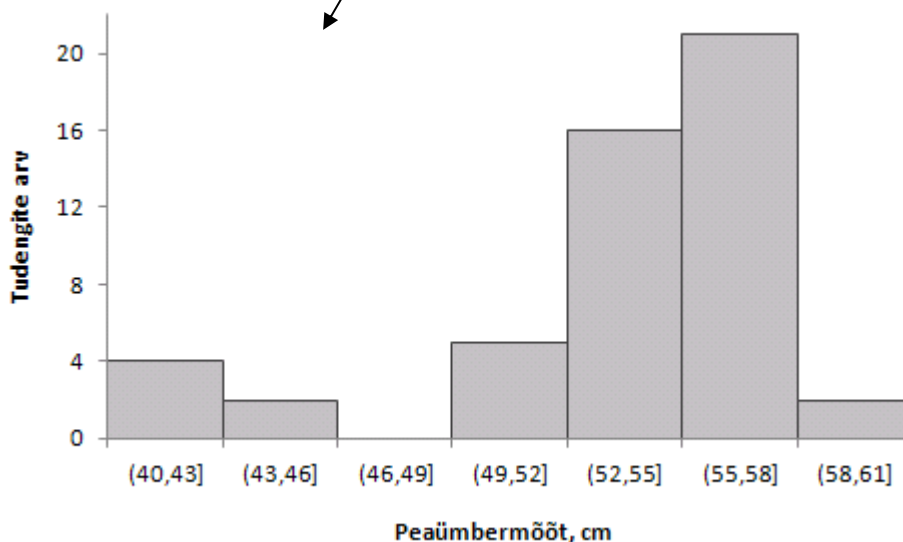
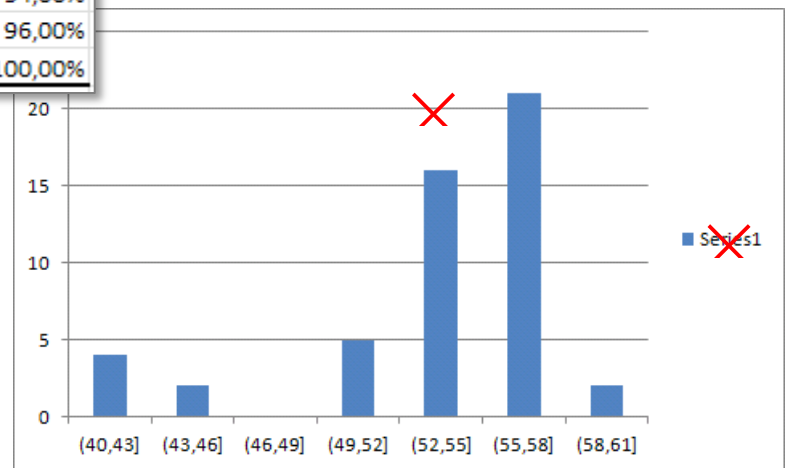
(**kumulatiivne sagedus** ehk **jaotus** näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

| X | Y | Z |
|-----------------|-----------|--------------|
| pea_ymb klassid | Frequency | Cumulative % |
| 43 | 4 | 8,00% |
| 46 | 2 | 12,00% |
| 49 | 0 | 12,00% |
| 52 | 5 | 22,00% |
| 55 | 16 | 54,00% |
| 58 | 21 | 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 % |
|---------|-----------------|-----------|--------------|
| (40,43] | 43 | 4 | 8,00% |
| (43,46] | 46 | 2 | 12,00% |
| (46,49] | 49 | 0 | 12,00% |
| (49,52] | 52 | 5 | 22,00% |
| (52,55] | 55 | 16 | 54,00% |
| (55,58] | 58 | 21 | 96,00% |
| (58,61] | 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 22 sammuga 4
- 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

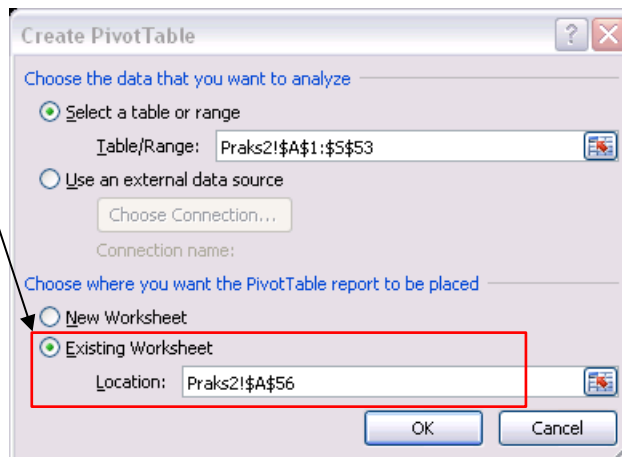
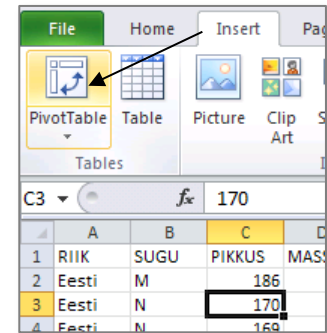
4. Sõnastage esimese ülesande lõpetuseks üks lause absoluutsete sageduste ja üks lause kumulatiivsete sageduste põhjal – pange need laused ka kirja.

Ülesanne 2.

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

Tööjuhend

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

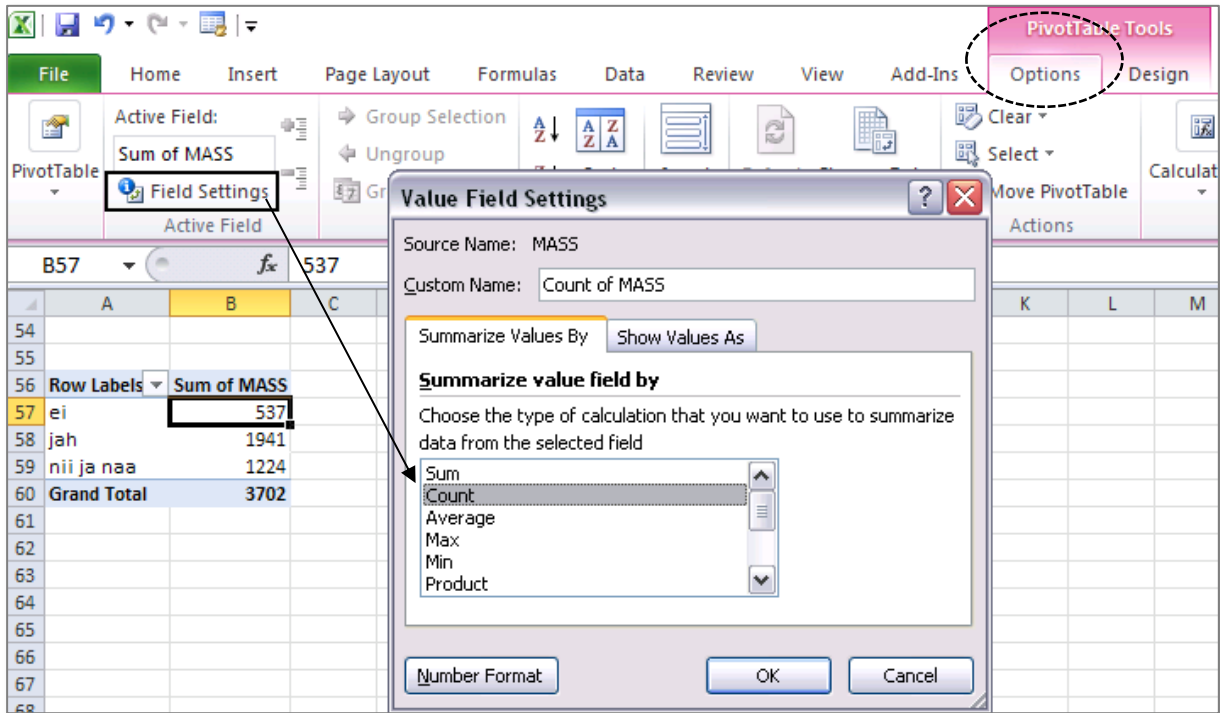


4.

Tulemuseks saadud tabel:

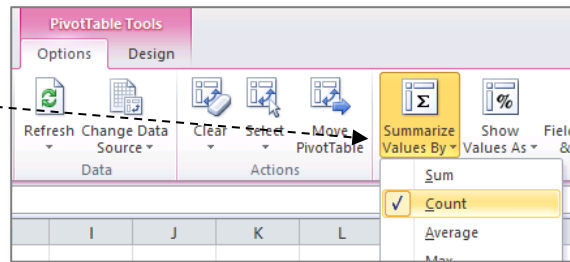
| Row Labels | Sum of MASS |
|--------------------|-------------|
| ei | 537 |
| jah | 1941 |
| nii ja naa | 1224 |
| Grand Total | 3702 |

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



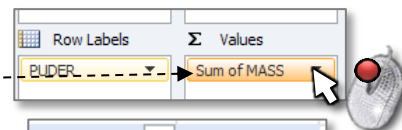
Selle osa võite vahele jätta

☞ **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'

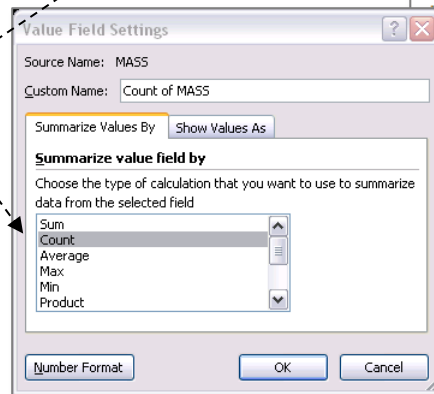


või **hiire parempoolse klahviga** *Pivot Table*'s

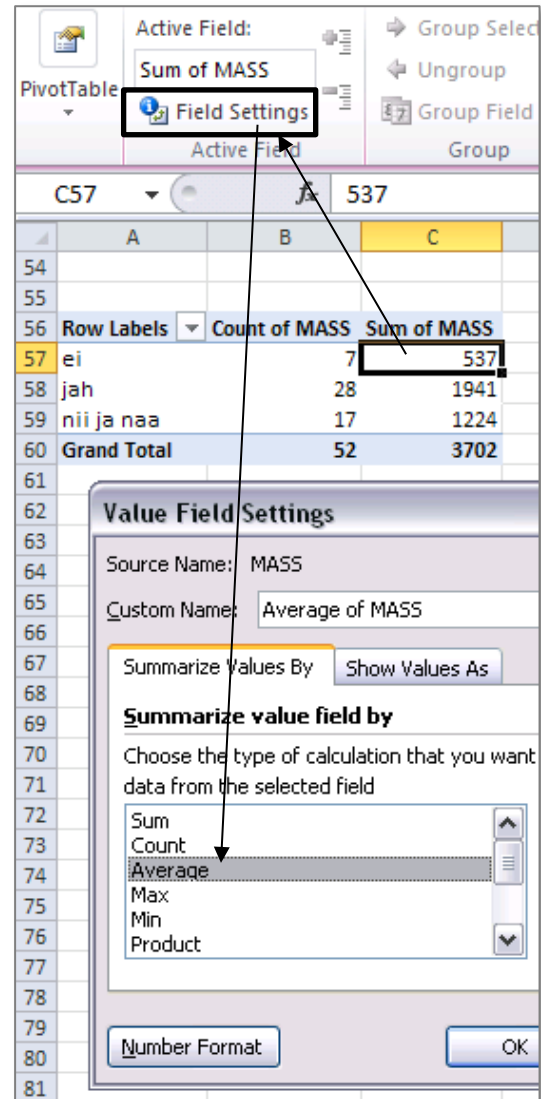
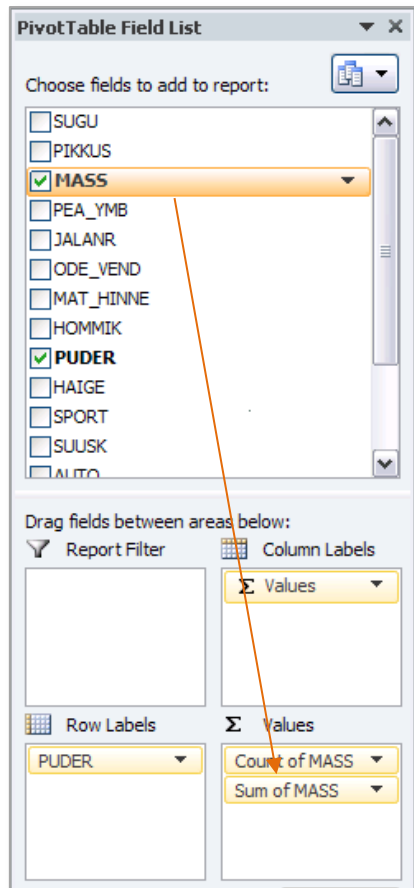
ning valides avanenud rippmenüüst käsu *Value Field Settings...* ja



sealt edasi vajaliku funktsiooni:



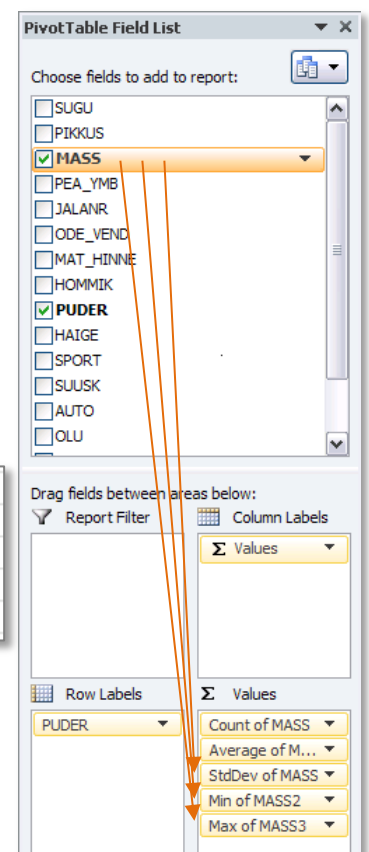
6. Arvutage lisaks ka keskmised kehamassid.



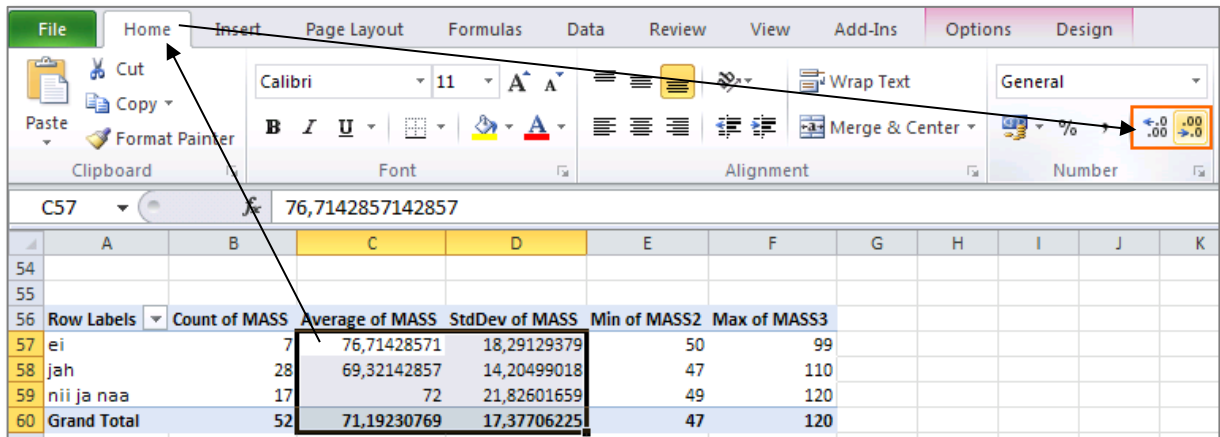
7. Arvutage ka kehamassi standardhälbed ning minimaalsed ja maksimaalsed väärtused.

Tulemus:

| Row Labels | Count of MASS | Average of MASS | StdDev of MASS | Min of MASS2 | Max of MASS3 |
|--------------------|---------------|--------------------|--------------------|--------------|--------------|
| ei | 7 | 76,71428571 | 18,29129379 | 50 | 99 |
| jah | 28 | 69,32142857 | 14,20499018 | 47 | 110 |
| nii ja naa | 17 | 72 | 21,82601659 | 49 | 120 |
| Grand Total | 52 | 71,19230769 | 17,37706225 | 47 | 120 |



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



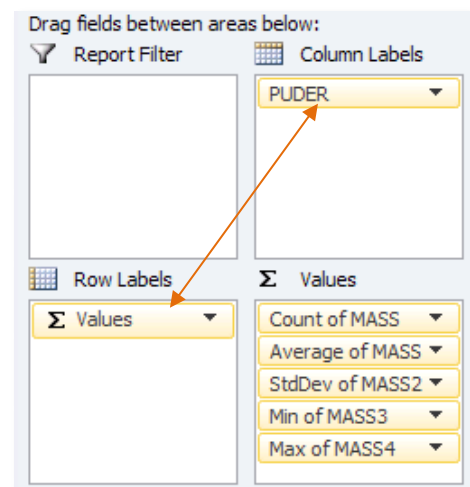
Tulemus:

| Row Labels | Count of MASS | Average of MASS | StdDev of MASS | Min of MASS2 | Max of MASS3 |
|--------------------|---------------|-----------------|----------------|--------------|--------------|
| ei | 7 | 76,7 | 18,3 | 50 | 99 |
| jah | 28 | 69,3 | 14,2 | 47 | 110 |
| nii ja naa | 17 | 72,0 | 21,8 | 49 | 120 |
| Grand Total | 52 | 71,2 | 17,4 | 47 | 120 |

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 | 28 | 17 | 52 |
| Average of MASS | 76,7 | 69,3 | 72,0 | 71,2 |
| StdDev of MASS | 18,3 | 14,2 | 21,8 | 17,4 |
| Min of MASS2 | 50 | 47 | 49 | 47 |
| Max of MASS3 | 99 | 110 | 120 | 120 |

Spikker:

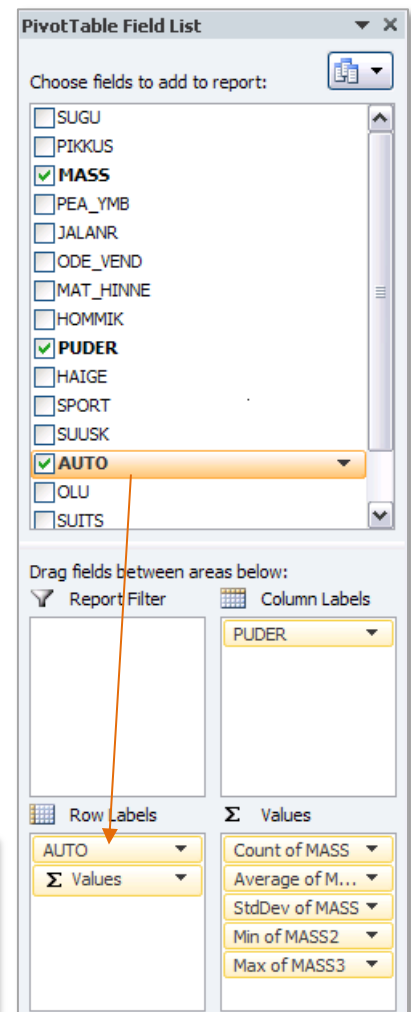


10. Ja nüüd jagage saadud tabel ridadeks auto omamist näitava tunnuse AUTO järgi.

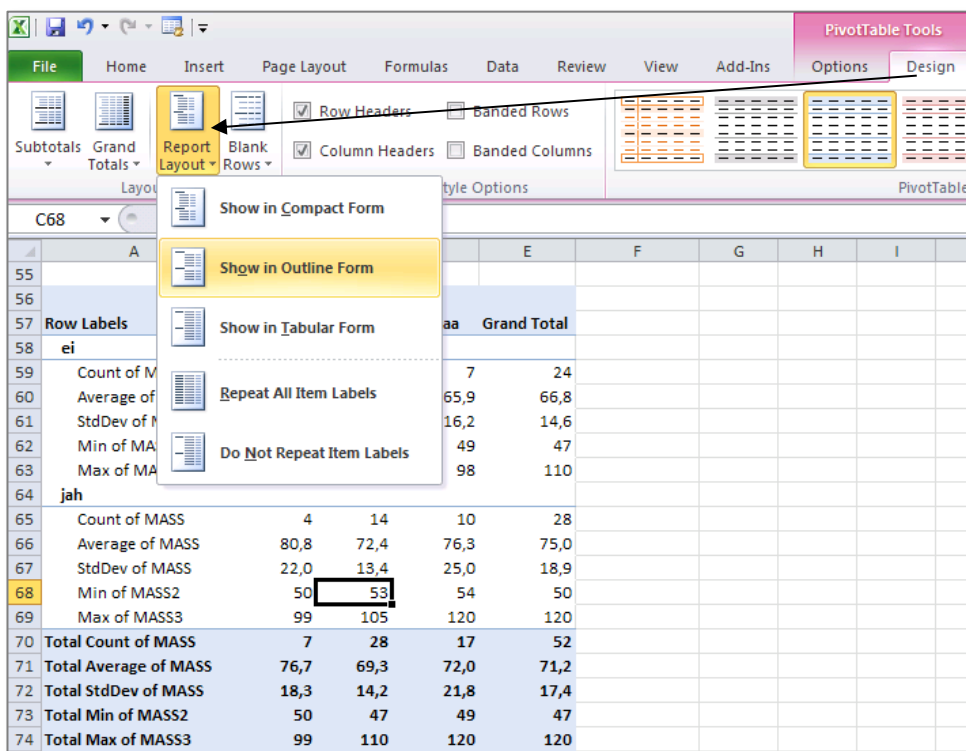
Oodatav tulemus:

| Row Labels | ei | jah | nii ja naa | Grand Total |
|------------------------------|-------------|-------------|-------------|-------------|
| ei | | | | |
| Count of MASS | 3 | 14 | 7 | 24 |
| Average of MASS | 71,3 | 66,2 | 65,9 | 66,8 |
| StdDev of MASS | 14,2 | 14,8 | 16,2 | 14,6 |
| Min of MASS2 | 55 | 47 | 49 | 47 |
| Max of MASS3 | 80 | 110 | 98 | 110 |
| jah | | | | |
| Count of MASS | 4 | 14 | 10 | 28 |
| Average of MASS | 80,8 | 72,4 | 76,3 | 75,0 |
| StdDev of MASS | 22,0 | 13,4 | 25,0 | 18,9 |
| Min of MASS2 | 50 | 53 | 54 | 50 |
| Max of MASS3 | 99 | 105 | 120 | 120 |
| Total Count of MASS | 7 | 28 | 17 | 52 |
| Total Average of MASS | 76,7 | 69,3 | 72,0 | 71,2 |
| Total StdDev of MASS | 18,3 | 14,2 | 21,8 | 17,4 |
| Total Min of MASS2 | 50 | 47 | 49 | 47 |
| Total Max of MASS3 | 99 | 110 | 120 | 120 |

Spikker:



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



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 mittedöövatel tudengitel.