

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 $n = 57$ (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaübermõõtude väärtused võiks jagada 7 või 8 klassi (sest $\sqrt{57} \approx 7,5$; samas ei ole see absoluutne tõde, kui klasside piiridega paremini sobib, võib väärtused jagada ka 5, 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 valite avanenud rippmenüüst teid huvitava funktsiooni (nt. *Min*),
 - c. valitud funktsiooni väärtus kuvatakse samas akna alumisel serval, aga seda ei trükita kuhugi tabeli lahtrisse,
 - d. mõne muu funktsiooni (näiteks *Max*) tarvis tuleb eelnevat korrata.

The screenshot shows an Excel spreadsheet with columns A through V. Column E is highlighted. A 'Customize Status Bar' menu is open, showing various options with checkboxes. The 'Minimum' and 'Maximum' options are highlighted with a red dashed box. The status bar at the bottom displays the following statistics: Average: 55,78947368, Count: 57, Sum: 3180. A mouse cursor is pointing at the 'Praks2' sheet tab.

- Maksimaalne peaümberrõõd tuleb 60 ja minimaalne 48 cm.
- Klasside intervallide leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovitatavate klasside arvuga ehk $(60-48)/7 \approx 1,71$ 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 47-49 cm.
- Kokkuvõttes võime tudengite peaümberrõõdudest moodustada seitse 2-sentimeetrist klassi:

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

Et *Excel* paigutab klassi piiriga võrdsed väärtused alumisse klassi, on korrektsem esitada klassid poollõikudena (sest kuidas muidu aru saada, kumba klassi näiteks 49 cm kuulub?):

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

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

| OLU | SUITS | TEATER | KINO | pea_ymb_klassid |
|-----|-------|------------|-----------------------|-----------------|
| 0 | ei | rohkem kui | viimase aasta jooksul | 49 |
| 0 | ei | rohkem kui | viimase aasta jooksul | 51 |
| 2 | ei | viimase a | viimase kuu jooksul | 53 |
| 0 | ei | viimase a | viimase kuu jooksul | 55 |
| 1 | ei | viimase a | viimase aasta jooksul | 57 |
| 0 | ei | viimase a | viimase aasta jooksul | 59 |
| 1 | ei | viimase k | viimase kuu jooksul | |
| 1 | ei | rohkem kui | viimase kuu jooksul | |

- 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 '61'= $(59,61]$ 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 59-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:

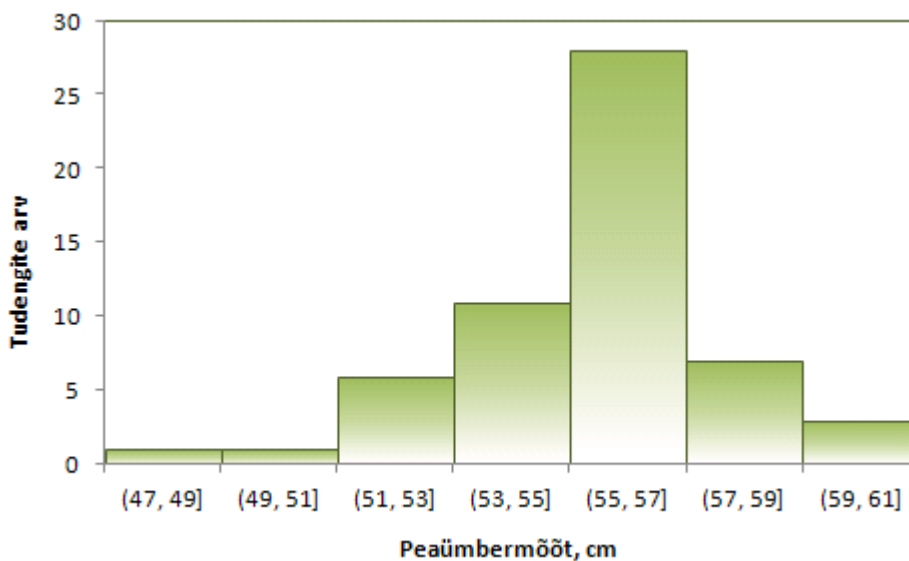
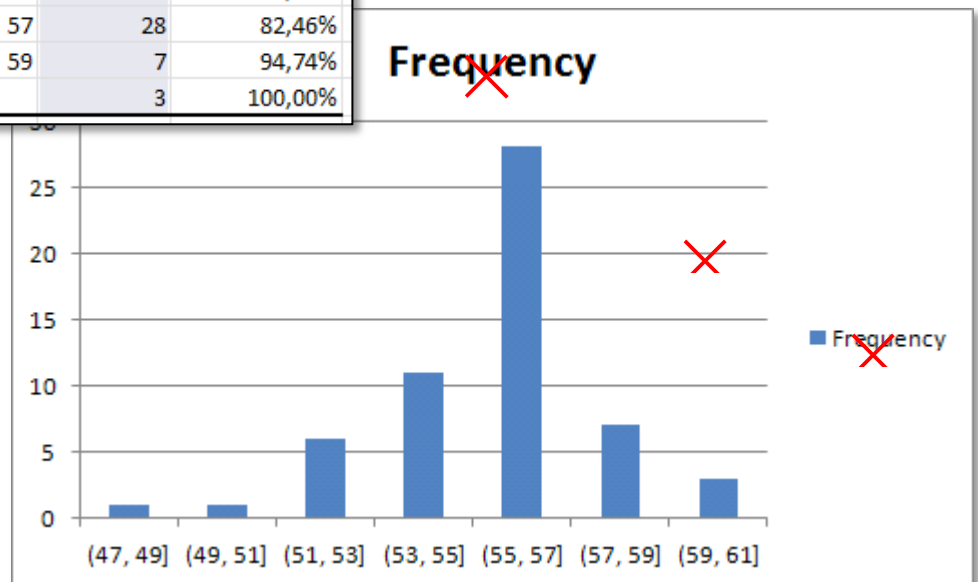
(kumulatiivne sagedus ehk jaotus näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

| X | Y | Z |
|-----------------|-----------|--------------|
| pea_ymb_klassid | Frequency | Cumulative % |
| 49 | 1 | 1,75% |
| 51 | 1 | 3,51% |
| 53 | 6 | 14,04% |
| 55 | 11 | 33,33% |
| 57 | 28 | 82,46% |
| 59 | 7 | 94,74% |
| More | 3 | 100,00% |

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

| Peaümberrööd | pea_ymb_klassid | Frequency | Cumulative % |
|--------------|-----------------|-----------|--------------|
| (47, 49] | 49 | 1 | 1,75% |
| (49, 51] | 51 | 1 | 3,51% |
| (51, 53] | 53 | 6 | 14,04% |
| (53, 55] | 55 | 11 | 33,33% |
| (55, 57] | 57 | 28 | 82,46% |
| (57, 59] | 59 | 7 | 94,74% |
| (59, 61] | More | 3 | 100,00% |

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



- Kustutage pealkiri, legend ja ruudujooned
- Kaotage ära hall kast joonise ümbert 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.

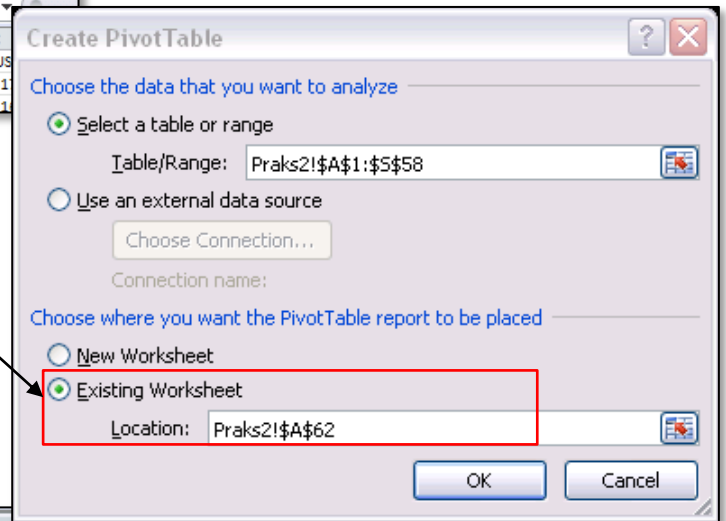
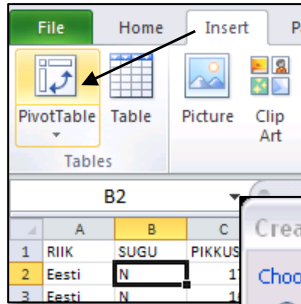
Ülesanne 2.

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.

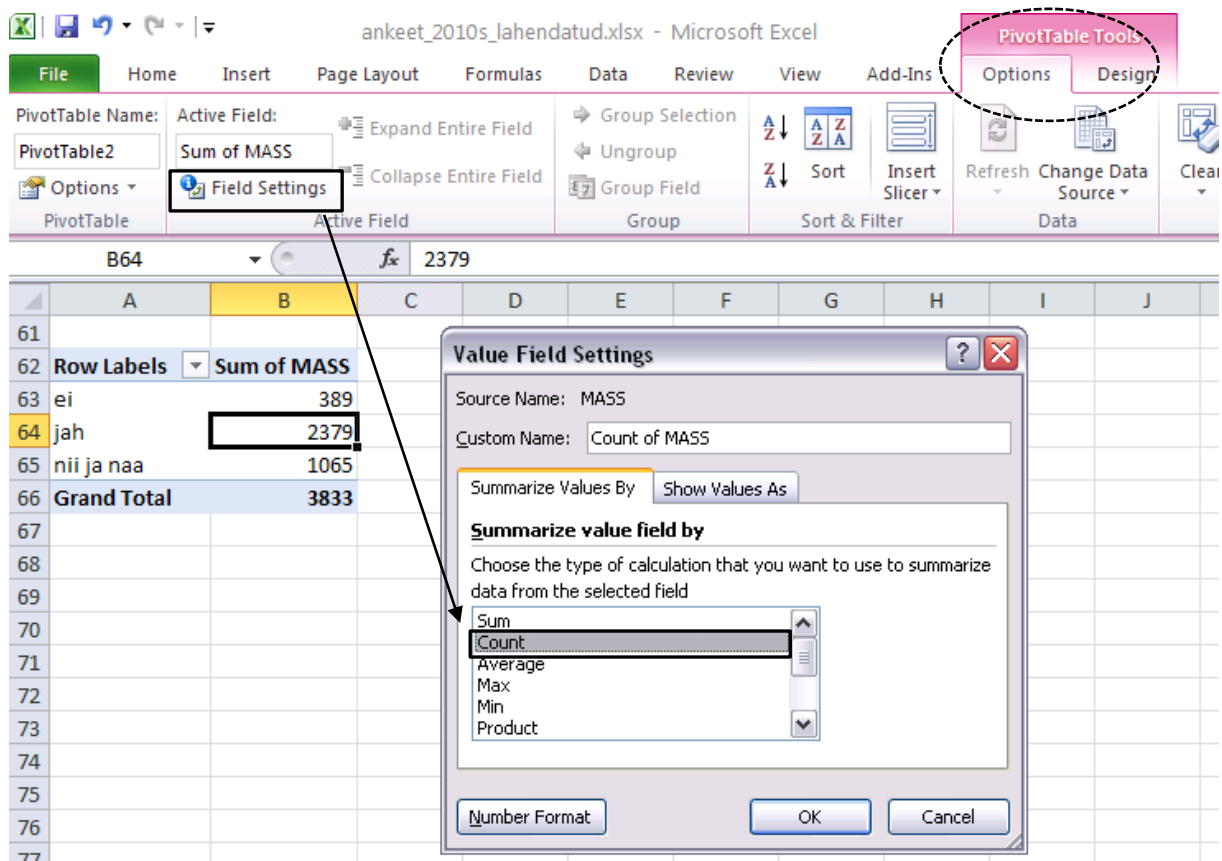
Lohistage tunnus PUDER lahtrisse Row Labels ning tunnus MASS lahtrisse Σ Values

Tulemus

Tulemuseks saadud tabel:

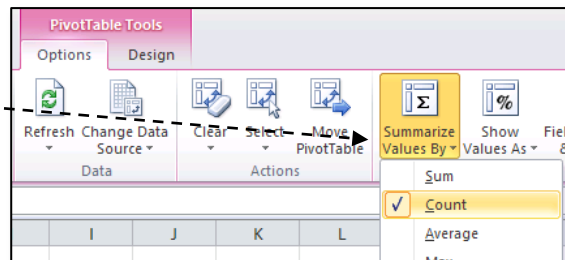
| Row Labels | Sum of MASS |
|--------------------|-------------|
| ei | 389 |
| jah | 2379 |
| nii ja naa | 1065 |
| Grand Total | 3833 |

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



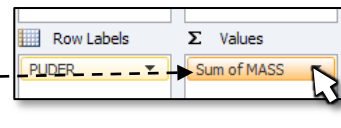
Selle osa võite vahele jätta

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



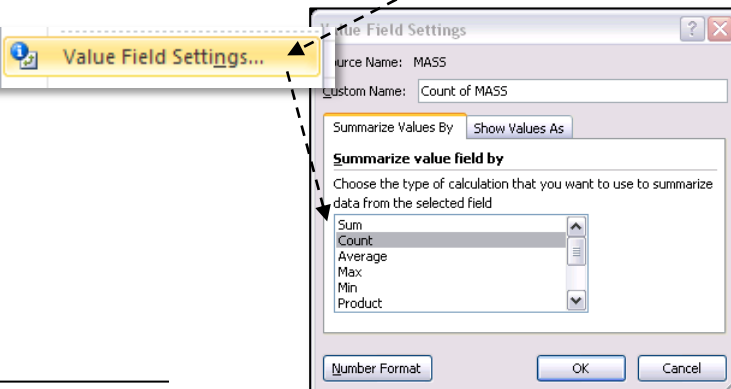
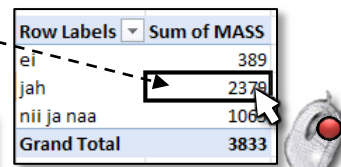
β) Rakendatavat funktsiooni saab muuta ka klikkides muudetaval tunnusel

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. In the PivotTable Field List, 'MASS' and 'PUUDER' are checked. The PivotTable shows 'Sum of MASS' for 'Count of MASS'. The Value Field Settings dialog is open, showing 'Average of MASS' as the summary calculation.

| Row Label | Count of MASS | Sum of MASS |
|--------------------|---------------|-------------|
| ei | 6 | 389 |
| jah | 35 | 2379 |
| nii ja naa | 16 | 1065 |
| Grand Total | 57 | 3833 |

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

Tulemus:

| Row Label | Count of MASS | Average of MASS | StdDev of MASS | Min of MASS | Max of MASS |
|--------------------|---------------|--------------------|--------------------|-------------|-------------|
| ei | 6 | 64,83333333 | 5,636192568 | 58 | 74 |
| jah | 35 | 67,97142857 | 24,99233496 | 49 | 200 |
| nii ja naa | 16 | 66,5625 | 7,667844112 | 53 | 80 |
| Grand Total | 57 | 67,24561404 | 19,97255322 | 49 | 200 |

The screenshot shows the PivotTable Field List with 'MASS' and 'PUUDER' selected. The PivotTable Field List shows 'Count of MASS', 'Average of MASS', 'StdDev of MASS', 'Min of MASS', and 'Max of MASS' selected.

- On saadud tulemused loomulikud? Maksimaalne kehamass 200 kg?

Põhimõtteliselt võimalik, aga antud juhul tundub pigem olevat tegu mõõtmis- või andmesisestusveaga.

- Taolised ilmselt vigased väärtused jäetakse tavaliselt analüüsist välja.

Antud juhul on kõige mõistlikum kehamassi väärtus 200 andmetabelist üles otsida ja lihtsalt ära kustutada.

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

| | | |
|-----|----------------|----|
| 180 | 65 | 58 |
| 183 | 200 | 59 |
| 175 | 60 | 56 |

→

| | | |
|-----|----|----|
| 180 | 65 | 58 |
| 183 | | 59 |
| 175 | 60 | 56 |

- Salvestage muudetud andmetabel peale tehtud muutust.

- Ühe sedavõrd erandliku väärtuse kustutamine muudab kõiki *Pivot Table*'i vastavasse ritta arvutatud arvarakteristikuid. Miks?

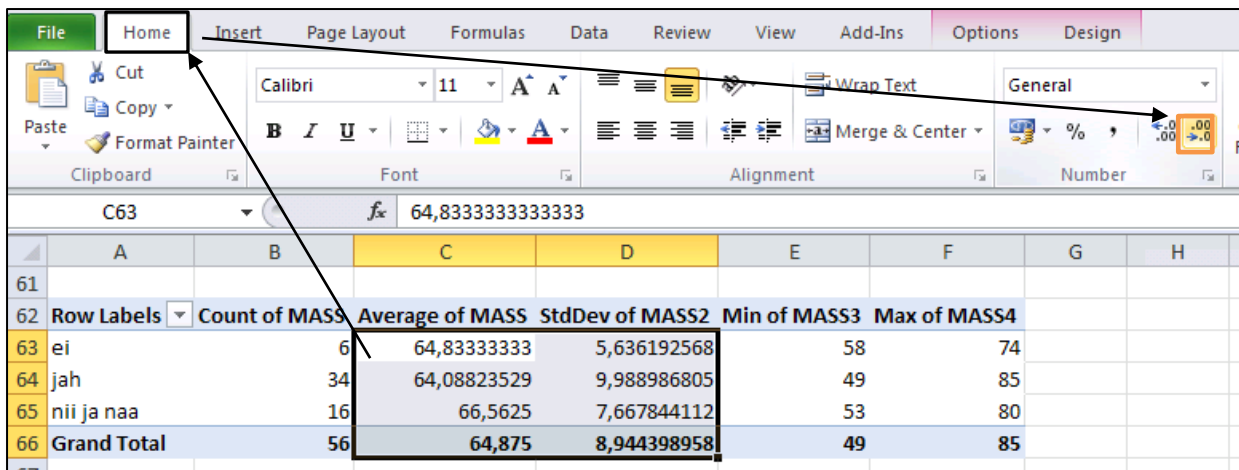
Et *Pivot Table*'s sisalduvad tulemused on kogu aeg seotud algandmete tabeliga, piisab kõigi arvarakteristikute muutunud andmetabeli alusel ümber arvutamiseks lihtsalt *Refresh*-käsu:

| Row Labels | Count of MASS | Average of MASS | StdDev of MASS2 | Min of MASS3 | Max of MASS4 |
|--------------------|---------------|--------------------|--------------------|--------------|--------------|
| ei | 6 | 64,83333333 | 5,636192568 | 58 | 74 |
| jah | 35 | 67,97142857 | 24,99233496 | 49 | 200 |
| nii ja naa | 16 | 66,5625 | 7,667844112 | 53 | 80 |
| Grand Total | 57 | 67,24561404 | 19,97255322 | 49 | 200 |

Tulemus:

| Row Labels | Count of MASS | Average of MASS | StdDev of MASS2 | Min of MASS3 | Max of MASS4 |
|--------------------|---------------|-----------------|--------------------|--------------|--------------|
| ei | 6 | 64,83333333 | 5,636192568 | 58 | 74 |
| jah | 34 | 64,08823529 | 9,988986805 | 49 | 85 |
| nii ja naa | 16 | 66,5625 | 7,667844112 | 53 | 80 |
| Grand Total | 56 | 64,875 | 8,944398958 | 49 | 85 |

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



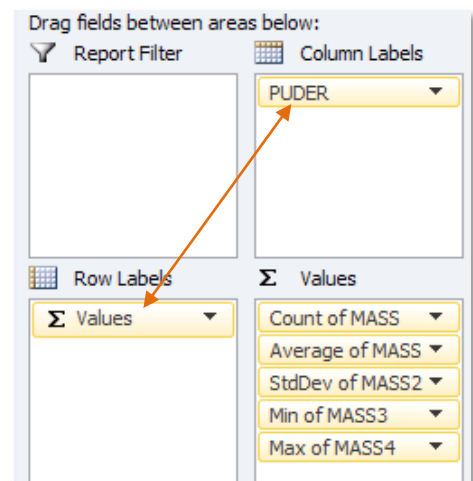
Tulemus:

| Row Labels | Count of MASS | Average of MASS | StdDev of MASS2 | Min of MASS3 | Max of MASS4 |
|--------------------|---------------|-----------------|-----------------|--------------|--------------|
| ei | 6 | 64,8 | 5,6 | 58 | 74 |
| jah | 34 | 64,1 | 10,0 | 49 | 85 |
| nii ja naa | 16 | 66,6 | 7,7 | 53 | 80 |
| Grand Total | 56 | 64,9 | 8,9 | 49 | 85 |

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 | 6 | 34 | 16 | 56 |
| Average of MASS | 64,8 | 64,1 | 66,6 | 64,9 |
| StdDev of MASS2 | 5,6 | 10,0 | 7,7 | 8,9 |
| Min of MASS3 | 58 | 49 | 53 | 49 |
| Max of MASS4 | 74 | 85 | 80 | 85 |

Spikker:

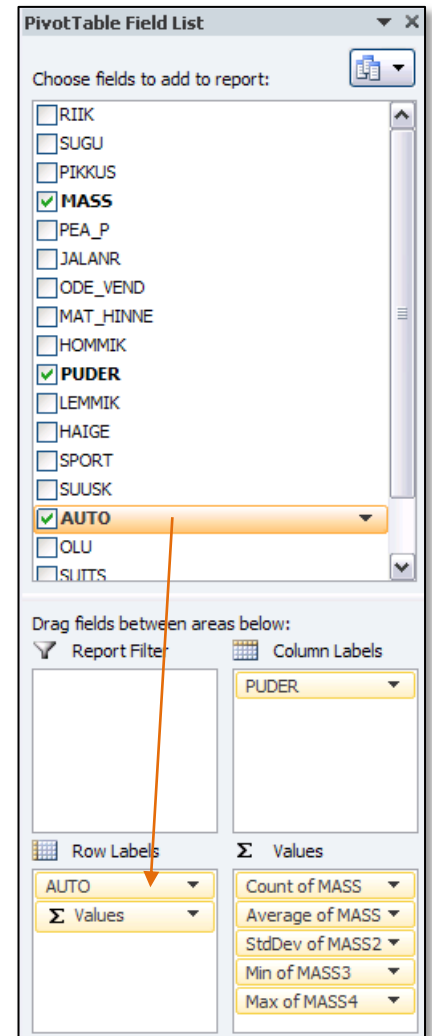


10. Ja nüüd jagage saadud tabel ridadeks ka veel auto omamist näitava tunnuse AUTO järgi.

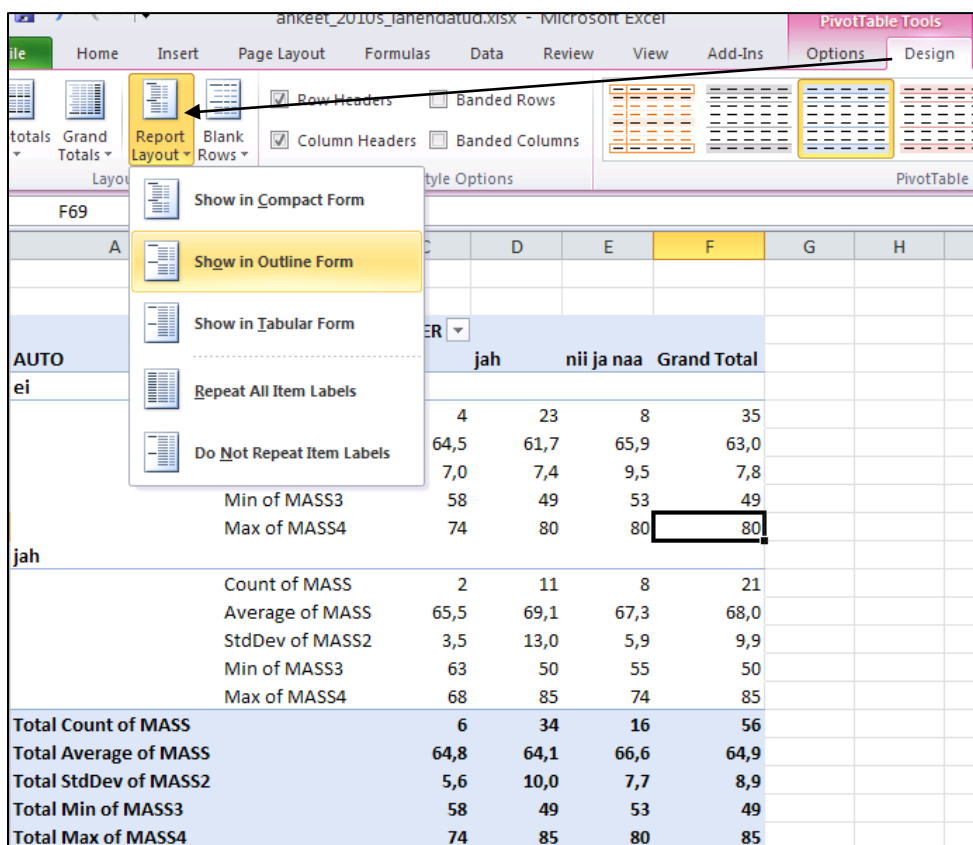
Oodatav tulemus:

| Row Labels | Column Label | | | Grand Total |
|------------------------------|--------------|-------------|-------------|-------------|
| | ei | jah | nii ja naa | |
| ei | | | | |
| Count of MASS | 4 | 23 | 8 | 35 |
| Average of MASS | 64,5 | 61,7 | 65,9 | 63,0 |
| StdDev of MASS2 | 7,0 | 7,4 | 9,5 | 7,8 |
| Min of MASS3 | 58 | 49 | 53 | 49 |
| Max of MASS4 | 74 | 80 | 80 | 80 |
| jah | | | | |
| Count of MASS | 2 | 11 | 8 | 21 |
| Average of MASS | 65,5 | 69,1 | 67,3 | 68,0 |
| StdDev of MASS2 | 3,5 | 13,0 | 5,9 | 9,9 |
| Min of MASS3 | 63 | 50 | 55 | 50 |
| Max of MASS4 | 68 | 85 | 74 | 85 |
| Total Count of MASS | 6 | 34 | 16 | 56 |
| Total Average of MASS | 64,8 | 64,1 | 66,6 | 64,9 |
| Total StdDev of MASS2 | 5,6 | 10,0 | 7,7 | 8,9 |
| Total Min of MASS3 | 58 | 49 | 53 | 49 |
| Total Max of MASS4 | 74 | 85 | 80 | 85 |

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
- kaerahelbeputru söövatel ja mitte söövatel tudengitel.