

## Biomeetria praks 2

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

1. Avage *MS Excelis* ankeedivastuste 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 'HEAD' (peaümberrõõtu) ja illustreerige saadud tabelit histogrammiga. Kujundage histogramm.

#### Tööjuhend

Pidevale arvtunnusele sagedustabeli konstrueerimiseks on *MS Excelis* vähemalt kolm võimalust – statistikaprotseduur *Histogram*, funktsioon *FREQUENCY* ja *PivotTable*. Neist kaks esimest 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  $n = 54$  (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaümberrõõtude väärtused võiks jagada 7 või 8 klassi (sest  $\sqrt{54} \approx 7,3$ ; 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 *Excelis* 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
1	GENDER	HEIGHT	WEIGHT	HEAD	JALANR
2	W	170	70	55.3	
3	W	158	47.5	55	
4	W	170	60	53	
5	W	170	50	55	
6	W	178	68	58	
7	W	163	56	40	
8	W	177	65	55	
9	W	162.5	53	55	
10	W	170	75	56	
11	M	175	74	57	
12	W	176	66	57	
13	M	175	64	56	
14	M	190	82	58	
15	W	161	50	55	
16	W	170	85	57	
17	W	176	58	52	
18	W	172	90	58	
19	W	158	55	57	
20	M	188	82	17	
21	W	169	60	55.3	
22	W	164	52	56	
23	W	172	62	56	
24	W	173	66	56	
25	W	169	60	55	
26	W	162	50	50	
27	W	165	52	50.3	
28	M	170	80	56	
29	M	176	74	56	
30	M	175	73	54	
31	W	171	63	57	
32	W	170	60	53	
33	W	163	62	55	
34	M	181	74	55	
35	W	168	60	55	
36	W	174	54	55	
37	W	166	68	56	
38	W	168	63	53	
39	W	165	58	56	
40	W	171	75	55	
41	W	165	77	58	
42	W	161	55	57	
43	M	183	75	75	
44	W	169	53	55	
45	W	175	60	57	
46	W	167	80	57.5	
47	W	158	70	55	
48	M	174	87	57	
49	W	165	61	57	
50	W	164	58	57	
51	W	185	80	60	
52	W	177	63	60	
53	W	160	70	57	
54	W	162	70	57	
55	W	172	58	57	

**Customize Status Bar**

- Cell Mode Ready
- Flash Fill Blank Cells
- Flash Fill Changed Cells
- 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 55.2037037
- Count 55
- Numerical Count 54
- Minimum 17
- Maximum 75
- Sum 2981
- Upload Status
- View Shortcuts
- Zoom Slider
- Zoom 56%

Andmed

AVERAGE: 55.2037037 COUNT: 55 NUMERICAL COUNT: 54 MIN: 17 MAX: 75 SUM: 2981

Maksimaalne peaümberrõõm tuleb 75 ja minimaalne 17 cm. Kas need arvud on reaalsed? Kui ümberrõõm on omale keeruline ette kujutada, siis diameetriga (lõõimõõduga) on ehk lihtsam. Siit ka küsimus – kui suur on pea diameeter, kui ümberrõõm on 17 cm, ja kas saadud tulemus on reaalne?

Spikker. Mõõletavasti avaldub ringi ümberrõõm valemiga  $P = 2\pi r$ , millest diameeter  $d = 2r = P/\pi$  (Excelit arvutusmasina kasutades võite sisestada valemikujul  $=17/PI()$ ).

Tulemuseks peaks tulema 5,4 cm. Võrdluseks – tavalise CD-plaadi lõõimõõm on 12 cm ja kohvitassil umbes 7 cm. Seega tundub peaümberrõõm 17 cm ilmselt liiga väike. Lisaks leidub tabelis peaümberrõõm 40 cm, mis on samuti pisut liiga väike (lõõimõõm 12,7 cm).

Samas on maksimaalne väärtus 75 cm peaümberrõõmduks jälle pisut liiga suur.

Ilmselt on nii 17, 40 kui ka 75 cm väärtuste puhul tegu mõõtmis- või andmete ülesmõõrkimisveaga.

Et traditsioonilised statistikameetodid ei ole mõeldud ebareaalsete või väga erandlike andmete analüüsimiseks, jäetakse sellised väärtused sageli analüüsist välja. Antud juhul on kõige mõistlikum peaümberrõõdu väärtused 17, 40 ja 75 cm andmetabelist lihtsalt ära kustutada.

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

- Minimaalne peaümberrõõd korrigeeritud andmetest on 50 ja maksimaalne 62 cm.
- Klasside intervallide ligikaudseks leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovitatavate klasside arvuga ehk  $(62-50)/7 = 1,7\dots$  cm.

Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli vajadusel veidi suurendada või vähendada, samuti võib korrigeerida intervallide arvu. Antud juhul võiks näiteks valida intervalli pikkuseks 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 peaümberrõõdudest moodustada seitse 2-sentimeetrist klassi:

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

Et *Excel* paigutab klassi piiriga võrdsed väärtused alumisse klassi, on korrektsem esitada klassid poollõikudena (sest kuidas muidu aru saada, kumba klassi näiteks 41 cm kuulub?):

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

NB. Soovi korral võinuks näiteks moodustada ka hoopis **ühemärgse 1,5-sentimeetrist** klassi kujul

(49;50,5], (50,5;52], (52;53,5], (53,5;55], (55;56,5], (56,5;58], (58;59,5], (59,5;61], (61;62,5];

või siis hoopis **seitse 2-sentimeetrist** klassi kujul

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

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

O	P	Q	
SMOKE		Head_upper_limits	
no		50	
no		52	
no		54	
no		56	
no		58	
no		60	
no			

- Märkusi:
  - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;

– *Excel* tõlgendab etteantud väärtusi järgmiselt:

'51' tähendab ' $\leq 51$ ';

'53' tähendab ' $51 < x \leq 53$ ' jne

(ehk igasse klassi kuuluvaks loetakse need väärtused, mis on väiksemad või võrdsed ette antud ülemisest klassi piirist ja mis ei kuulu eelnevasse klassidesse);

- **viimast klassi '62'=(60,62] ette ei anta**, sest sagedustabeli moodustamisel teeb *Excel* ise täiendava klassi, kuhu loeb kokku kõik eelnevasse klassidesse mittekuuluvad väärtused. Seega jäävad täiendavasse automaatselt loodavasse klassi kõik 60-st suuremad väärtused.

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

The screenshot shows the Microsoft Excel interface with the **DATA** tab selected. The **Data Analysis** tool is open, and the **Histogram** option is selected. The **Histogram** dialog box is displayed with the following settings:

- Input Range:** \$D\$1:\$D\$55
- Bin Range:** \$Q\$1:\$Q\$7
- Labels:**
- Output options:**
  - Output Range:** \$T\$1
  - Cumulative Percentage:**

The background spreadsheet shows a table with columns: C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T. The data in column D (HEAD) is highlighted in red. The data in column Q (Head\_upper\_limits) is also highlighted in red. The data in column T is highlighted in green.

Andmed  
Klassipiirid

Ütleb Excelile, 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)

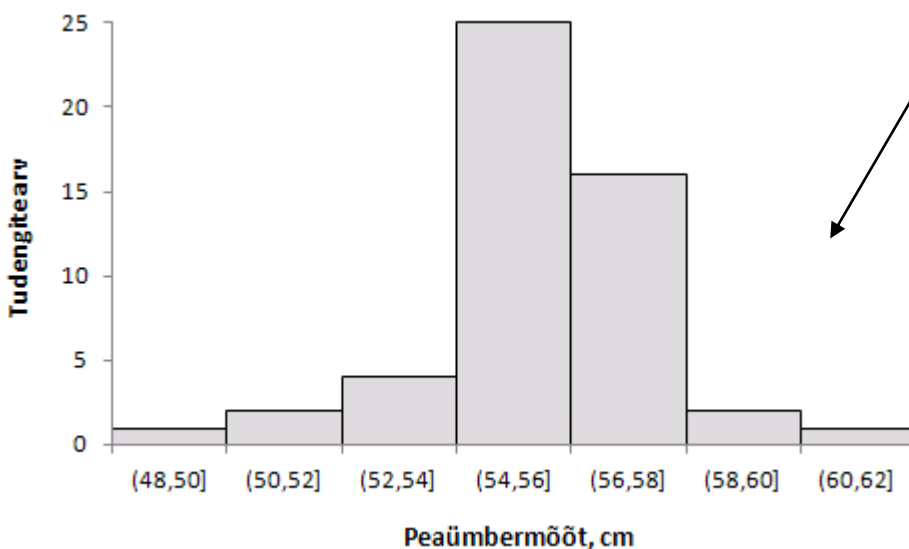
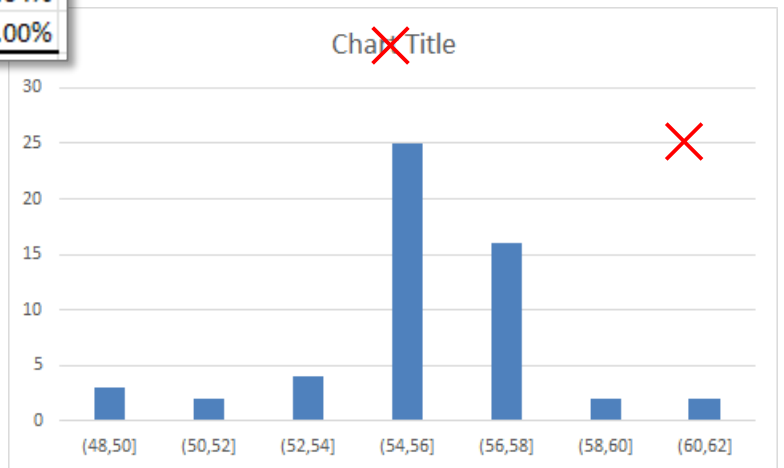
T	U	V
Head_upper_lir	Frequency	Cumulative %
50	1	1.96%
52	2	5.88%
54	4	13.73%
56	25	62.75%
58	16	94.12%
60	2	98.04%
More	1	100.00%

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

Head_upper_lir	Frequency	Cumulative %
(48,50]	1	1.96%
(50,52]	2	5.88%
(52,54]	4	13.73%
(54,56]	25	62.75%
(56,58]	16	94.12%
(58,60]	2	98.04%
(60,62]	1	100.00%

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

- Kustutage joonise pealkiri ja ruudujooned
- Kaotage ära hall kast joonise ümbert
- Lisage teljed ja määrake y-telje ulatuseks 0 kuni 25 sammuga 5
- Lisage pealkirjad x- ja y-teljele
- Tulpade vahe võtke nulliks
- Esitage telgede ühikud ja nimed kirjasuuruses 10 ning musta värvi
- Tulbad värvige nii, nagu enesele meeldib, lisage tulpadele ka sisust veidi tumedamad piirjooned



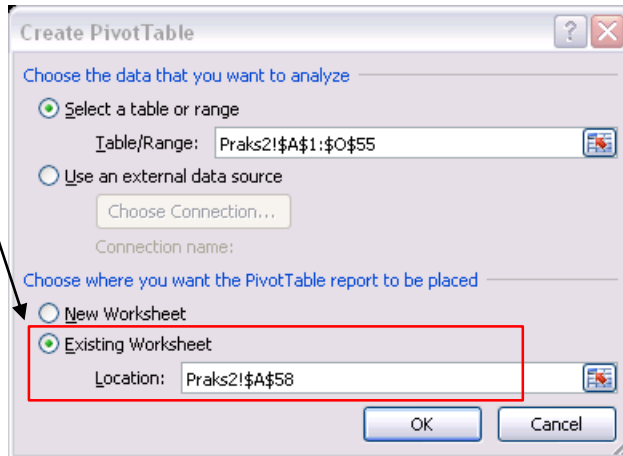
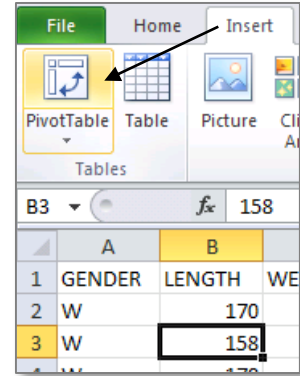
4. Sõnastage esimese ülesande lõpetuseks üks lause absoluutsete sageduste ja üks lause kumulatiivsete sageduste põhjal – pange need laused ka kirja.

## Ülesanne 2.

Konstrueerige *PivotTable*'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:

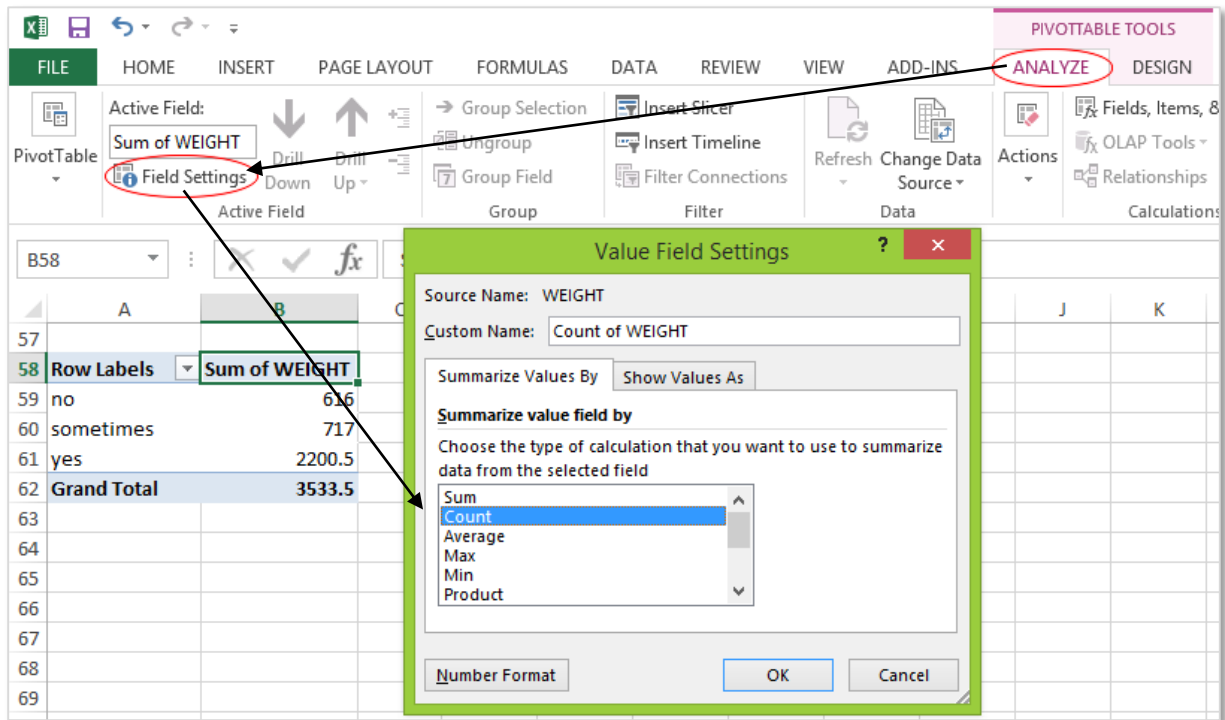


4.

Tulemuseks saadud tabel:

Row Labels	Sum of WEIGHT
no	616
sometimes	717
yes	2200.5
<b>Grand Total</b>	<b>3533.5</b>

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



6. Arvutage lisaks ka keskmised kehamassid.

The image shows the process of configuring a PivotTable in Microsoft Excel. On the left, the **PivotTable Field List** task pane is open, showing a list of fields. The **WEIGHT** and **PORRIDGE** fields are checked. The **WEIGHT** field is highlighted, and an orange arrow points from it to the **Value Field Settings** dialog box. The **PivotTable** is displayed in the background, with the **Field Settings** button in the ribbon highlighted. The **Value Field Settings** dialog box is open, showing the **Source Name** as **WEIGHT** and the **Custom Name** as **Average of WEIGHT**. The **Summarize Values By** section is set to **Show Values As**, and the **Summarize value field by** section is set to **Average**. The **Number Format** button is also visible.

Row Labels	Count of WEIGHT	Sum of WEIGHT
no	9	616
sometimes	11	717
yes	34	2200.5
<b>Grand Total</b>	<b>54</b>	<b>3533.5</b>



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

Tulemus:

Row Labels	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT:	Max of WEIGHT
no	9	68.44444444	11.95942213	50	85
sometimes	11	65.18181818	11.60015674	52	87
yes	34	64.72058824	10.2033797	47.5	90
<b>Grand Total</b>	<b>54</b>	<b>65.43518519</b>	<b>10.66206744</b>	<b>47.5</b>	<b>90</b>

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

Row Labels	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT:	Max of WEIGHT
no	9	68.44444444	11.95942213	50	85
sometimes	11	65.18181818	11.60015674	52	87
yes	34	64.72058824	10.2033797	47.5	90
<b>Grand Total</b>	<b>54</b>	<b>65.43518519</b>	<b>10.66206744</b>	<b>47.5</b>	<b>90</b>

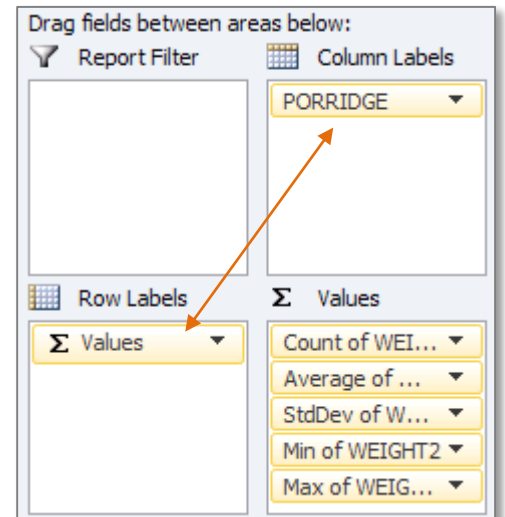
Tulemus:

Row Labels	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT:	Max of WEIGHT
no	9	68.4	12.0	50	85
sometimes	11	65.2	11.6	52	87
yes	34	64.7	10.2	47.5	90
<b>Grand Total</b>	<b>54</b>	<b>65.4</b>	<b>10.7</b>	<b>47.5</b>	<b>90</b>

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

Values	Column Labels			Grand Total
	no	sometimes	yes	
Count of WEIGHT	9	11	34	54
Average of WEIGHT	68.4	65.2	64.7	65.4
StdDev of WEIGHT	12.0	11.6	10.2	10.7
Min of WEIGHT2	50	52	47.5	47.5
Max of WEIGHT3	85	87	90	90

Spikker:

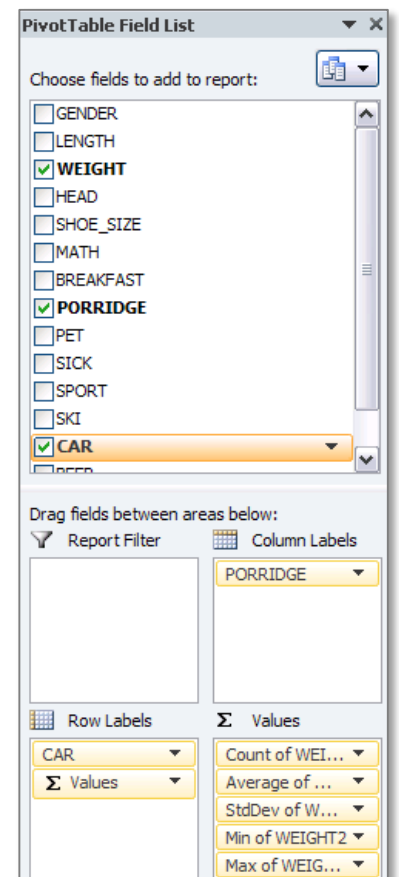


10. Ja nüüd jagage saadud tabel ridadeks auto omamist näitava tunnuse CAR järgi ning jätke tabelist välja üks tudeng, kes ei tea, kas tal on auto või mitte.

Oodatav tulemus:

Row Labels	Column Labels			Grand Total
	no	sometimes	yes	
<b>no</b>				
Count of WEIGHT	4	5	13	22
Average of WEIGHT	69.0	62.8	63.6	64.4
StdDev of WEIGHT	11.7	8.0	10.4	9.9
Min of WEIGHT2	58	52	50	50
Max of WEIGHT3	85	73	90	90
<b>yes</b>				
Count of WEIGHT	5	5	21	31
Average of WEIGHT	68.0	70.2	65.4	66.6
StdDev of WEIGHT	13.5	13.8	10.3	11.1
Min of WEIGHT2	50	53	47.5	47.5
Max of WEIGHT3	82	87	82	87
<b>Total Count of WEIGHT</b>	<b>9</b>	<b>10</b>	<b>34</b>	<b>53</b>
<b>Total Average of WEIGHT</b>	<b>68.4</b>	<b>66.5</b>	<b>64.7</b>	<b>65.7</b>
<b>Total StdDev of WEIGHT</b>	<b>12.0</b>	<b>11.3</b>	<b>10.2</b>	<b>10.6</b>
<b>Total Min of WEIGHT2</b>	<b>50</b>	<b>52</b>	<b>47.5</b>	<b>47.5</b>
<b>Total Max of WEIGHT3</b>	<b>85</b>	<b>87</b>	<b>90</b>	<b>90</b>

Spikker:



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

The screenshot shows the Microsoft Excel interface with the PivotTable Tools ribbon active. The 'DESIGN' tab is selected. The 'Report Layout' dropdown menu is open, showing options for how to display the PivotTable. The background shows a PivotTable with columns for 'sometimes yes' and 'Grand Total'.

		sometimes yes	Grand Total
Count of WEIGHT		5	13
Average of WEIGHT		62.8	63.6
StdDev of WEIGHT		8.0	10.4
Min of WEIGHT3	58	52	50
Max of WEIGHT4	85	73	90
yes			
Count of WEIGHT	5	5	21
Average of WEIGHT	68.0	70.2	65.4
StdDev of WEIGHT2	13.5	13.8	10.3
Min of WEIGHT3	50	53	47.5
Max of WEIGHT4	82	87	82
Total Count of WEIGHT	9	10	34
Total Average of WEIGHT	68.4	66.5	64.7
Total StdDev of WEIGHT2	12.0	11.3	10.2
Total Min of WEIGHT3	50	52	47.5
Total Max of WEIGHT4	85	87	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
- putru söövatel ja mittedöövatel tudengitel.