

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 lehekülje 'Praks2' ülemisse vasakusse nurka.

Ülesanne 1.

Konstrueerige sagedustabel tunnusele 'PEA_P' (peaümberrõõd) ja illustreerige saadud tabelit histogrammiga. Kujundage histogramm.

Tööjuhend

Pidevale arvtunnusele sagedustabeli tegemiseks on *MS Excel*'is kaks moodust – statistikaprotseduur *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 seeläbi ka 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 $n = 47$ (Saite ka sellise arvu? Kuidas?) annab vihje, et tudengite peaümberrõõdude väärtused võiks jagada $\sqrt{47} \approx 7$ klassi (see ei ole absoluutne tõde, kui klasside piiridega paremini sobib, võib väärtused jagada ka 5, 6 või hoopis 8 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 (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.

	A	B	C	D	E	F	G	H	I	J	K
1	RIIK	SUGU	PIKKUS	MASS	PEA P	JALANR	ODE_V	EMAT_HIN	HOMMIK	PUDER	LEMM
2	Eesti	N	170	64	57	38	1	4	helbed võ	jah	jah
3	Eesti	N	170	60	57	40	0	4	võileib	nii ja naa	jah
4	Eesti	N	163	64	54	38	2	5	helbed võ	jah	jah
5	Eesti	N	161,5	52	54	37,5	2	4	puder	jah	jah
6	Eesti	N	171	75	55,5	40	1	4	puder	jah	jah
7	Soome	N	178	70	52	40	1	5	puder	jah	jah
8	Eesti	N	175	67	57	40	2	5	helbed võ	nii ja naa	jah
9	Eesti	N	160	52	56	38	2	4	võileib	nii ja naa	jah
10	Eesti	N	160	55	56	36	3	4	võileib	jah	jah
11	Eesti	N	170	69	52	36	1	4	võileib	nii ja naa	jah
12	Eesti	N	167	57	57	38,5	1	4	ei söö tav	ei	ei
13	Eesti	N	172	59	55,5	38	1	5	ei söö tav	nii ja naa	jah
14	Eesti	N	167	60	46	39	1	3	puder	jah	jah
15	Eesti	N	175	63	57	40	3	3	võileib	jah	jah
16	Eesti	N	173	73	52	41	0	4	võileib	nii ja naa	ei
17	Eesti	N	168	65	56	39	1	5	võileib	ei	jah
18	Soome	N	170	80	57,5	39	2	3	võileib	jah	jah
19	Eesti	N	165	70	56	39	4	3	võileib	jah	jah
20	Soome	N	170	60	55,5	38	2	3	puder	jah	jah
21	Eesti	N	175	60	55	38	0	4	võileib	jah	jah
22	Eesti	N	160	50	56	37	2	4	muu	nii ja naa	jah
23	Eesti	N	160	51	56	36	1	4	võileib	jah	jah
24	Eesti	N	167	55	56	37	2	4	puder	jah	jah
25	Eesti	N	161	54	54	38	2	4	võileib	jah	ei
26	Eesti	N	165	54,3	54	38	2	4	ei söö tav	ei	jah
27	Eesti	N	158	55	55	35	0	4	ei söö tav	nii ja naa	jah
28	Soome	N	173	62	54	39	1	5	võileib	jah	jah
29	Eesti	N	174	58	54	39	2	5	muu	jah	jah
30	Eesti	M	175	76	57	43	2	3	võileib	jah	jah
31	Eesti	N	180	65	55	40	1	4	puder	jah	jah
32	Eesti	N	178	64	57	40	0	5	helbed võ	jah	jah
33	Soome	N	164	63	42	38	2	4	võileib	jah	jah
34	Soome	N	172	70	43,5	39	2	4	võileib	ei	jah
35	Eesti	N	161	56	55	38	1	4	puder	jah	jah
36	Eesti	N	181	60	55	40	1	4	võileib	jah	jah
37	Eesti	N	164	55	53	38	1	4	puder	jah	jah
38	Eesti	N	177	71	57	41	2	4	võileib	nii ja naa	jah
39	Eesti	N	175	70	54	41	1	4	võileib	nii ja naa	jah
40	Eesti	N	170	110	57	43	1	4	muu	nii ja naa	jah
41	Eesti	N	164	58	54	38,5	1	4	võileib	ei	ei
42	Eesti	N	163	50,5	55,5	37	1	4	võileib	jah	ei
43	Eesti	N	159	57	56	37	1	4	võileib	ei	jah
44	Eesti	N	163	54	52	38	1	4	helbed võ	ei	jah
45	Eesti	N	175	55	55	38,5	1	4	võileib	ei	jah
46	Eesti	N	160	70	55	39	1	4	võileib	jah	jah
47	Eesti	M	185	82	51	45	1	4	muu	jah	jah
48	Eesti	N	174	60	57	39	1	4	puder	jah	jah
49											
50											

- Maksimaalne peaümberrõõm tuleb 59 ja minimaalne 42 cm.
 - Klasside intervallide leidmiseks tuleb maksimaalsest väärtusest lahutada minimaalne ning jagada see soovivate klasside arvuga ehk $(59-42)/7=2,43$ cm.
- Et klasside piirid oleks „ümmargused“ tuleb leitud intervalli veidi suurendada või vähendada, antud ülesande puhul sobiks intervalli pikkuseks ilmselt 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 peaümberrõõmudest 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 58 cm kuulub?):

$$(40,43], (43,46], (46,49], (49,52], (52,55], (55,58] \text{ ja } (58,61].$$

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

OLU	SUITS	TEATER	KINO	pea_ymb_klassid
0	enam ei, a	viimase aa	viimase 10 päeva jooksul	43
1,5	jah	viimase aa	viimase 10 päeva jooksul	46
0	ei	viimase aa	viimase kuu jooksul	49
0	ei	viimase aa	viimase kuu jooksul	52
1,5	jah	viimase aa	viimase kuu jooksul	55
0	ei	viimase aa	viimase 10 päeva jooksul	58
1	ei	viimase aa	viimase 10 päeva jooksul	

- Märkusi:
 - mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – miks?;
 - *Excel* tõlgendab ette antud 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 klassi 58-st suuremad väärtused.

2. Sagedustabel protseduuri *Histogram* abil:

Tools / Tööriistad → Data Analysis... → Histogram

The screenshot shows the Excel interface with the **Data Analysis** task pane open. The **Histogram** option is selected. The **Histogram** dialog box is displayed with the following settings:

- Input Range:** \$E\$1:\$E\$48
- Bin Range:** \$U\$1:\$U\$7
- Labels**
- Output options:**
 - Output Range:** \$X\$1
 - New Worksheet Ply:**
 - New Workbook**
 - Pareto (sorted histogram)**
 - Cumulative Percentage**
 - Chart Output**

Red dashed arrows indicate the flow from the **Data Analysis** dialog to the **Histogram** dialog, and from the **Histogram** dialog to the **Input Range** and **Bin Range** fields. A red dashed arrow also points from the **Labels** checkbox to the **Labels** text box below.

Ütleb Excel'ile, et ette antud lahtri-
blokkide esimeses reas on nimed

Väljundtabeli (vasaku ülemise nurga)
asukoht

Lisaks tavalistele sagedustele
arvutatakse ka kumulatiivsed
suhtelised sagedused e. jaotus

Protseduuri *Histogram* oodatav tulemus:

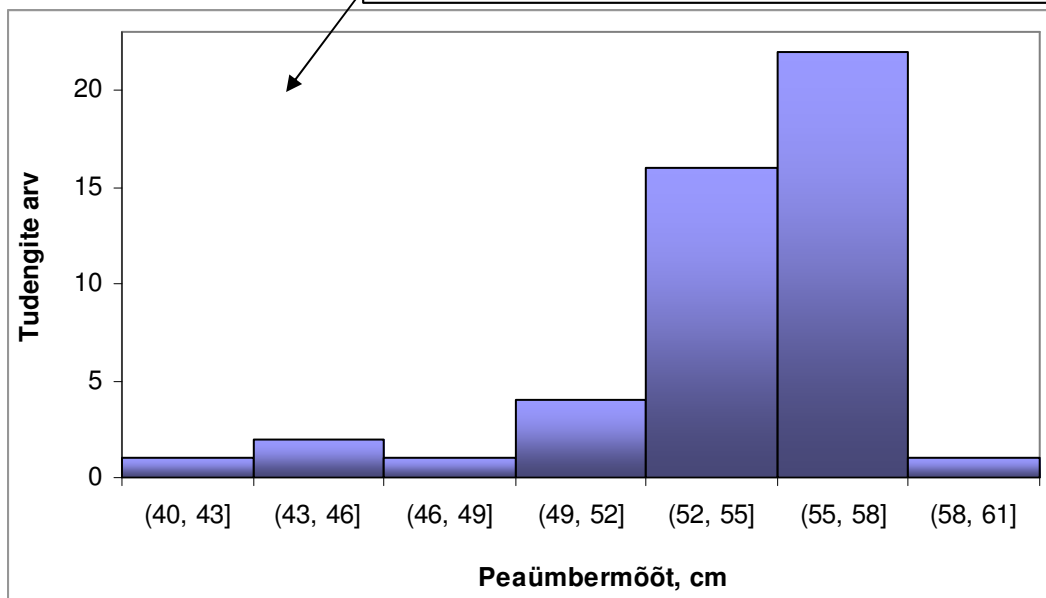
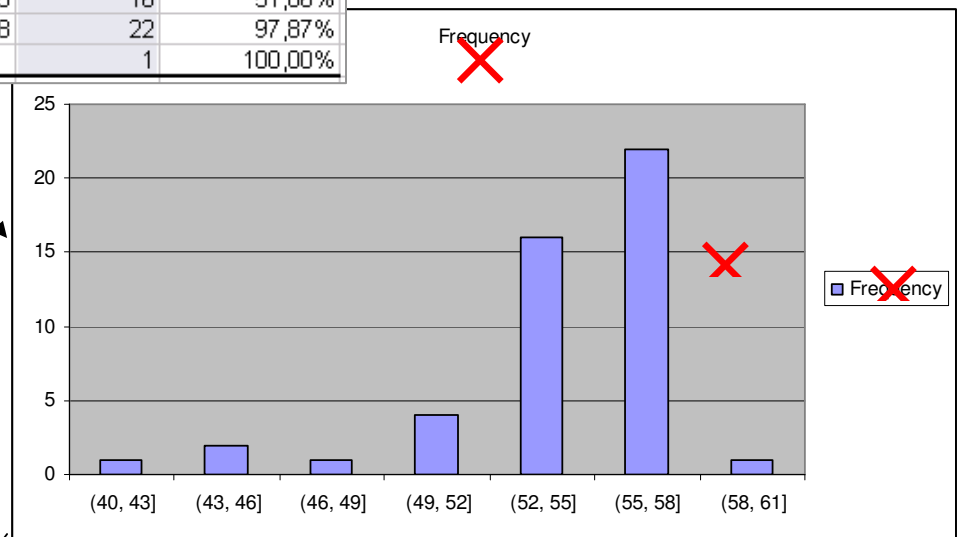
(kumulatiivne sagedus e jaotus näitab, kui mitu % vaatlustest on väiksemad või võrdsed vastava klassi ülemisest piirist)

X	Y	Z
pea ymb klassid	Frequency	Cumulative %
43	1	2,13%
46	2	6,38%
49	1	8,51%
52	4	17,02%
55	16	51,06%
58	22	97,87%
More	1	100,00%

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

Peaümbermõõt	pea ymb klassid	Frequency	Cumulative %
(40, 43]	43	1	2,13%
(43, 46]	46	2	6,38%
(46, 49]	49	1	8,51%
(49, 52]	52	4	17,02%
(52, 55]	55	16	51,06%
(55, 58]	58	22	97,87%
(58, 61]	More	1	100,00%

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



- Pealkiri ja legend ✗
- Taust valgeks
- Ruudujooned ✗
- Pealkirjad x- ja y-teljetele
- Tulpade vahe = 0
- Telgede ühikud ja nimed kirjasuuruses 10
- Y-telje maksimum = 23 (ja miinimum = 0)

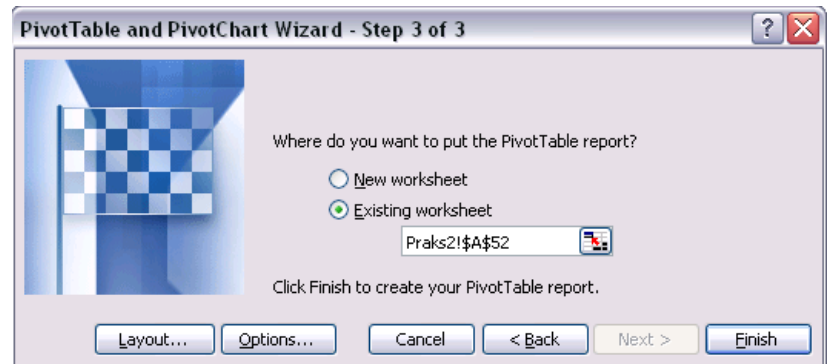
4. Sõnastage üks lause absoluutsete sageduste ja üks lause kumulatiivsete sageduste põhjal.

Ülesanne 2.

Konstrueerige tabel, kirjeldamaks tudengite kehamassi sõltuvalt kaerahelbepudru söömisest, kasutades *Pivot Table*'t.

Tööjuhend

1. Paigutage kursor andmetabeli suvalisse lahtrisse.
2. *Data / Andmed* → *PivotTable and PivotChart Report... / PivotTable ja PivotChart Aruanne...*
3. Loodav tabel paigutage 'Praks2'-lehele andmetabeli alla:



4.

Sum of MASS	
PUDER	Total
ei	470,3
jah	1754,5
nii ja naa	736
Grand Total	2960,8

5. Leidke *Exceli* poolt vaikumisi arvatud kehamasside summa asemel tudengite arv erinevates ridades.

The screenshot shows the 'PivotTable Field' dialog box with 'Source field: MASS' and 'Name: Count of MASS'. Under 'Summarize by:', 'Count' is selected. The PivotTable below shows the following data:

PUDER	Total
ei	8
jah	28
nii ja naa	11
Grand Total	47

6. Arvutage lisaks ka keskmised kehamassid.

The screenshot shows the 'PivotTable Field List' with 'MASS' selected. The PivotTable below shows the following data:

PUDER	Data	Total
ei	Count of MASS	8
	Sum of MASS	470,3
jah	Count of MASS	28
	Sum of MASS	1754,5
nii ja naa	Count of MASS	11
	Sum of MASS	736
Total Count of MASS		47
Total Sum of MASS		

The 'PivotTable Field' dialog box shows 'Source field: MASS' and 'Name: Average of MASS'. Under 'Summarize by:', 'Average' is selected.

PUDER	Data	Total
ei	Count of MASS	8
	Average of MASS	58,7875
jah	Count of MASS	28
	Average of MASS	62,66071429
nii ja naa	Count of MASS	11
	Average of MASS	66,90909091
Total Count of MASS		47
Total Average of MASS		62,99574468

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

PUDER	Data	Total
ei	Count of MASS	8
	Average of MASS	58,7875
	StdDev of MASS	5,716251894
	Min of MASS2	54
	Max of MASS3	70
jah	Count of MASS	28
	Average of MASS	62,66071429
	StdDev of MASS	8,450349074
	Min of MASS2	50,5
	Max of MASS3	82
nii ja naa	Count of MASS	11
	Average of MASS	66,90909091
	StdDev of MASS	16,38569221
	Min of MASS2	50
	Max of MASS3	110
Total Count of MASS		47
Total Average of MASS		62,99574468
Total StdDev of MASS		10,58602355
Total Min of MASS2		50
Total Max of MASS3		110

8. Esitage viimati konstrueeritud tabel kujul, kus igale arvutatud suurusele vastab üks veerg ja igale mannapudru söömise väärtusele üks rida.

Ümardage keskmised ja standardhälbed 1 kohani peale koma.

PUDER	Data	Total
ei	Count of MASS	8
	Average of MASS	58,7875
	StdDev of MASS	5,716251894

PUDER	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	8	58,7875	5,716251894	54	70
jah	28	62,66071429	8,450349074	50,5	82
nii ja naa	11	66,90909091	16,38569221	50	110
Grand Total	47	62,99574468	10,58602355	50	110

PUDER	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	8	58,8	5,7	54	70
jah	28	62,7	8,5	50,5	82
nii ja naa	11	66,9	16,4	50	110
Grand Total	47	63,0	10,6	50	110

9. Oskate te saadud tabelit viia tagasi punktis 7 näidatud kujule?

Aga järgnevale kujule?

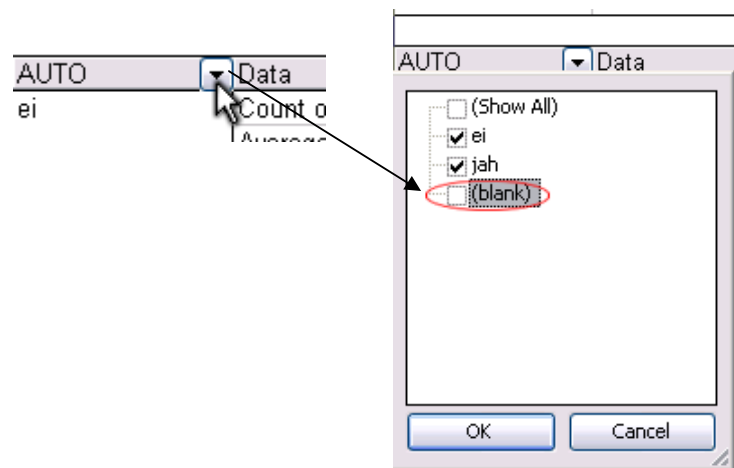
	PUDER			
Data	ei	jah	nii ja naa	Grand Total
Count of MASS	8	28	11	47
Average of MASS	58,8	62,7	66,9	63,0
StdDev of MASS	5,7	8,5	16,4	10,6
Min of MASS2	54	50,5	50	50
Max of MASS3	70	82	110	110

Aga järgnevale kujule?

		PUDER			
AUTO	Data	ei	jah	nii ja naa	Grand Total
ei	Count of MASS	6	12	8	26
	Average of MASS	59,7	61,5	64,3	61,9
	StdDev of MASS	6,4	6,7	8,5	7,1
	Min of MASS2	54	52	50	50
	Max of MASS3	70	76	73	76
jah	Count of MASS	2	15	3	20
	Average of MASS	56,0	63,8	74,0	64,5
	StdDev of MASS	1,4	10,0	31,4	14,1
	Min of MASS2	55	50,5	52	50,5
	Max of MASS3	57	82	110	110
(blank)	Count of MASS		1		1
	Average of MASS		60,0		60,0
	StdDev of MASS		#DIV/0!		#DIV/0!
	Min of MASS2		60		60
	Max of MASS3		60		60
Total Count of MASS		8	28	11	47
Total Average of MASS		58,8	62,7	66,9	63,0
Total StdDev of MASS		5,7	8,5	16,4	10,6
Total Min of MASS2		54	50,5	50	50
Total Max of MASS3		70	82	110	110

Ühel tudengil on küsimusele auto olemasolu kohta vastamata (*Excel* tekitab puuduva väärtuse tarvis grupi '(blank)' e '(tühi)').

Jätke vastavad read tabelist välja.



		PUDER			
AUTO	Data	ei	jah	nii ja naa	Grand Total
ei	Count of MASS	6	12	8	26
	Average of MASS	59,7	61,5	64,3	61,9
	StdDev of MASS	6,4	6,7	8,5	7,1
	Min of MASS2	54	52	50	50
	Max of MASS3	70	76	73	76
jah	Count of MASS	2	15	3	20
	Average of MASS	56,0	63,8	74,0	64,5
	StdDev of MASS	1,4	10,0	31,4	14,1
	Min of MASS2	55	50,5	52	50,5
	Max of MASS3	57	82	110	110
Total Count of MASS		8	27	11	46
Total Average of MASS		58,8	62,8	66,9	63,1
Total StdDev of MASS		5,7	8,6	16,4	10,7
Total Min of MASS2		54	50,5	50	50
Total Max of MASS3		70	82	110	110

10. 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.