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' (peaümbermõõt) ja illustreerige saadud tabelit histogrammiga. Kujundage histogramm.

Tööjuhend

Pidevale arvtunnusele sagedustabeli tegemiseks on *MS Excel*'is kaks moodust – funktsioon *FREQUENCY* ja statistikaprotseduur *Histogram*. 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 44 (Saite ka sellise numbri? Kuidas?) annab vihje, et tudengite peaümbermõõtude väärtused võiks jagada √44≈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 annavad minimaalne ja maksimaalne väärtus. Nende leidmiseks on *Excel*'is päris palju erinevaid mooduseid; kõige kiirem variant neid numbreid teada saada, ilma neid kuhugi töölehele välja arvutamata on järgmine:
 - a. võtate blokki huvipakkuvad lahtrid,
 - b. klikite hiire <u>parempoolse</u> 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	0	8_	E	F	G	Н	1	J	
1	SUGU	PIKKUS	MASS	PEA P	MAT_HINNE	EBA_AINE	AINEKOOD	PUDER	ÖNNELI	к	
2	N	165	59		4	füüsika	Reaal	jah	jah		
3	N	165	49		4	kirjandus	Hum	jah	jah		
4	N	171	70	56	3	ajalugu	Hum	ei	jah		
5	M	186	75		4	eesti keel	Hum	jah	jah		
6	M	182	73	56	4	keemia	Reaal	nii ja naa	jah		
7	N	172,5	60		4	füüsika	Reaal	jah	jah		
8	N	167	75	Sec. 1	4			ei	jah		
9	N	175	64	50	3	inglise keel	Hum	jah	jah		
10	N	169	72	55	3	eesti keel	Hum	jah	jah	3	
11	M	174	93	58	3	matemaatika	Reaal	jah	jah		
12	M	191	88	60	3	vene keel	Hum	jah	nii ja naa	3	
13	M	175	70	56	5	vene keel	Hum	iah	nii ia naa	3	
14	N	157	57		5	geograafia	Hum	ei	iah		
15	M	187	70	56	3	nrantsuse keel	Hum	iah	iah		-
16	M	187	56		3	inglise keel	Hum	iah	iah		-
17	N	164	50		4	matemaatika	Reaal	iah	iab		-
18	M	185	80	56	3	matemaatika	Rezal	oi	iab		-
10	M	170	90	56	4	füücika	Reaal	nii ia naa	pii ja pas		-
20	NI NI	177	70	52	2	füüsika	Poppi	nii ja naa	nii ja naa		
20	M	101	00	60	3	kiriandus	Hum	ei	nii ja naa	-	-
21	IVI N	181	1.0	64	4	(ibiekeene?)	Hum	ei	nii ja naa	3	
22	N	102	1,0	04	3	uniskonnaop.	Hum	el	nii ja naa	1	-
23	M	188	18		4	muusika ajal.	Hum	el	nii ja naa	3	-
24	N	168	00	50	3	TUUSIKA	Keaal	jah	jah		-
25	N	175	70	56	6	ajalugu	Hum	jah	jah		
26	M	186	80	66	3	keemia	Reaal	jah	jah		-
27	N	169	55		3	matemaatika	Reaal	jah	ei		
28	N	171	54	53,2	3	keemia	Reaal	ei	jah		-
29	N	178	80	56	5	keemia	Reaal	nii ja naa	jah	_	
30	N	157	48	Same	4	füüsika	Reaal	jah	jah		
31	N	180	67	61	5	soome keel	Hum	ei	jah		
32	M	190	85		3	keemia	Reaal	jah	nii ja naa	3	
33	N	173	70		4	matemaatika	Reaal	ei	jah		
34	M	168	123	43	5	kehaline kasv.	Muu	jah	ei		
35	N	175	70		3	matemaatika	Reaal	ei	jah		
36	M	181	69	51	3	keemia	Reaal	jah	jah		
37	M	195	87	58	3	matemaatika	Reaal	jah	jah		
38	M	186	95	50	3	vene keel	Hum	jah	nii ja naa	3	
39	N	158	42		4	keemia	Reaal	ei	jah		
40	M	183	80	200	3	matemaatika	Reaal	jah	nii ja naa	3	
41	M	179,5	78	52	3	saksa keel	Hum	ei	ei		
42	M	183	85	55	3	keemia	Reaal	ei	nii ia naa	3	
43	N	164	58		4	bioloogia	Hum	iah	iah		-
44	M	177	71		3	keemia	Reaal	iah	iah	3	
45	M	177	75	56	4	eesti keel	Hum	iah	nii ia naa		
46	N	165	46		3	eesti keel	Hum	iah	nii ia naz	3	
47	M	176	80	52	4	saksa keel	Hum	nii ia naa	ei		-
48	M	185	100	58	3	matemaatika	Read	ei	iah		-
40	hd	177	75		3	ajalugu	Hum	iah	iab		-
50	N	177	62	56	3	kehaline kasv	hâm	ei	iab		1
51	M	186	76	55	3	keemia	Rezal	iab	iab		
52	M	179	69	57	4	vene kaal	Hum	jan	jah		-
52	M	172	72	50	4	vene keel	Hum	jan pii ja pee	jan pii ia car		-
54	M	185	50	56	2	vene keel	Hum	ini ja rida jak	nii ja naa		-
40	N N	100	08	50	3	matomastik-	Rozal	jdli jek	im ja 1182	•	-
00	NI NI	107	00		4	maternaatika	Doord	jan iet	jan jan		-
50	N N	101	00	50	4	keemia	reaai	jan	jan		-
50	N	101	44	00	3	ajalugu	Hum Dec-1	el	jan j=t		
08	N	109	OU	23	4	tuusika	Keaal	jan	jah		-
59	N	175	01		3	geograafia	Hum	jah	nii ja naa	3	
60	N	158	50	55	4	vene keel	Hum	ei	jah		-
61	N	167	62,5	55	3	füüsika	Reaal	jah	jah		
62	M	186	95	57	3	matemaatika	Mar	ne	h		
63	N	170	60	55	3	füüsika	NO	10	h	_	
64	M	197	85	50	3	matemaatika			h		-
65	N	153	65	57	4	ajalugu	<u>H</u> V6	erage	h		
66	N	170	60	58	3	keemia	- C	un t	h		-
67	N	170	73	57,5	4	s7' el	<u>_</u> 01	1110	i		
68						C CAR	- C	under Dile voor o			
69							COL	une Nums	,		
								~			
70						STREET, STREET	Ma:	~			
70						1991 1		-			
70 71 72							Mi-				
70 71 72 70				1/0	Lat Dra		Min	Ê.		141	
70 71 72	1 >	▶\\ Ar	idmec	d / Pra	aks1 ∖Pra	ks2/	Min	i 		<	
70 71 72	↓ 	M \ Ar	idmec	d / Pra	aks1 ∖Pra	ks2/	Min Sur	n n		Kin-2	2

 Maksimaalne peaümbermõõt tuleb 200 ja minimaalne 23 cm. Kas need arvud on reaalsed? Kui ümbermõõtu on omale keeruline ette kujutada, siis diameetriga (läbimõõduga) on ehk lihtsam. Siit ka küsimus – kui suur on pea diameeter, kui ümbermõõt on 23 cm, ja kas see on reaalne?

<u>Spikker</u>. Mäletatavasti avaldub ringi ümbermõõt valemiga $P = 2\pi r$, millest diameeter $d = 2r = P/\pi$ (*Excelit* arvutusmasina kasutades võite sisestada valemi kujul '=23/PI()'). Tulemuseks peaks tulema ~7,3 cm. See on sama palju, kui tavalisel kohvitassil. Seega tundub peaümbermõõt 23 cm pisut liiga väike.

Ka suuruselt järgmine peaümbermõõt, 43 cm, ei ole kuigi reaalne (pea diameetriks tuleb 13,7 cm – võrdluseks, CD-plaadi läbimõõt on näiteks 12 cm).

Peaümbermõõt 200 cm on aga ilmselgelt liiga suur (kui suur see diameetrer siis tuleb?). Ilmselt on kõigil juhtudel tegu mõõtmis- või andmete ülesmärkimisveaga. Et **traditsioonilised statistikameetodid ei ole mõeldud ebareaalsete või väga erandlike andmete analüüsimiseks**, jäetakse sellised väärtused tavaliselt analüüsidest välja. Antud juhul on kõige mõistlikum peaümbermõõdud väärtustega 23, 43 ja 200 cm andmetabelist lihtsalt ära kustutada.

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

44	56	3	44	56	5
50	23	4	50		4
51		3 →	51		3

- Minimaalne peaümbermõõt korrigeeritud andmetest on 50 ja maksimaalne 66 cm.
- Teadmise, et 41 tudengi peaümbermõõdud varieeruvad 50-st 66 sentimeetrini, alusel võiks moodustada kuus 3-sentimeetrist klassi:

49-52; 52-55; 55-58; 58-61; 61-64; 64-67.

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?):

(49,52], (52,55], (55,58], (58,61], (61,64] ja (64,67].

 Moodustatud klasside *Excel*'ile ette andmiseks tuleb sisestada <u>klasside ülemisi piire</u> sisaldav abitabel:

AINE	AINEKOOD	PUDER	ŐNNELIK	Pea_ymb_klassid	Τ
ka	Reaal	jah	jah	52	Τ
dus	Hum	jah	jah	55	Τ
gu	Hum	ei	jah	58	Τ
keel	Hum	jah	jah	61	Ι
nia	Reaal	nii ja naa	jah	64	
ka	Reaal	jah	jah		

Märkusi:

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

- Excel tõlgendab ette antud väärtusi järgmiselt:

'52' tähendab '≤52';

'55' tähendab '52<x≤55' jne

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

– viimast klassi '67'=(64,67] ette ei anta, sest sagedustabeli moodustamisel teeb *Excel* ise täiendava klassi, kuhu loeb kokku kõik eelnevatesse klassidesse mitte kuuluvad väärtused. Seega jäävad täiendavasse klassi 64-st suuremad väärtused.

2. Sagedustabel protseduuri *Histogram* abil:



Tools / Tööriistad → Data Analysis...→ Histogram

Tanel Kaart, Mirjam Vallas, Snežana Ševtsova, Alo Tänavots

100,00%

1

Protseduuri Histogram oodatav tulemus:	N	0	• ► P	
(l	Pea_ymb_klassid	Frequency	Cumulative %	
(kumulatiivne sagedus e jaotus naitab,	52	7	17,07%	
kui mitu % vaatlustest on väiksemad	55	9	39,02%	
või võrdsed vastava klassi ülemisest piirist)	58	21	90,24%	
- .	61	3	97,56%	
	64	0	97.56%	

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

More



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

Ülesanne 2.

Konstrueerige tabel, kirjeldamaks tudengite kehamassi sõltuvalt mannapudru 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:





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



6. Arvutage lisaks ka keskmised kehamassid.



Tanel Kaart, Mirjam Vallas, Snežana Ševtsova, Alo Tänavots

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

PUDER [🖌 Data 🖉 🔽	Total
ei	Count of MASS	22
]	Average of MASS	67,39090909
]	StdDev of MASS	21,15228028
	Min of MASS2	1,6
	Max of MASS3	100
jah	Count of MASS	39
	Average of MASS	69,06410256
	StdDev of MASS	15,68530132
	Min of MASS2	46
	Max of MASS3	123
nii ja naa	Count of MASS	5
	Average of MASS	78,8
	StdDev of MASS	6,833739825
	Min of MASS2	72
	Max of MASS3	89
Total Count of	MASS	66
Total Average of	of MASS	69,24393939
Total StdDev of	f MASS	17,30499375
Total Min of MA	ASS2	1,6
Total Max of M	ASS3	123

Laske silmad saadud tulemustest üle.

Kas kõik tundub normaalne (vt minimaalseid ja maksimaalseid väärtusi)?

Total StdDev of MASS	17,30499375
Total Min of MASS2	1,6
Total Max of MASS3	123

Kehamass 1,6 kg?

Ilmselgelt on selle väärtuse näol tegu veaga. Ja vigaste väärtustega ei ole teha muud, kui need andmetabelist ära kustutada (kui muidugi ei ole võimalik kusagilt korrektset väärtust järele vaadata).



Kehamass 123 kg?

Võimalik. Aga otsides vastava arvu andmetabelist üles, ilmneb, et samas reas on pikkuseks märgitud 168 cm ja peaümbermõõt, mis sai küll juba eelmises ülesandes ära kustutatud, oli 43 cm. Need väärtused, eriti taolises kombinatsioonis, on enam kui kummalised.

Et antud juhul ei ole tegu mingi tõsise teadusuuringuga (viimase korral oleks ilmtingimata vajalik nende väärtuste päritolu täpsemalt uurida), siis on mõistlik kahtlusi äratanud arvud tabelist lihtsalt ära kustutada.



Ja kustutage need kaks kehamassi väärtust ning üks pikkuse väärtus ära ka 'Andmete'lehel paiknevast tabelist.

PS. Reaalseis teadusuuringuis kustutatakse sageli ära terve rida, sest kui kui juba osa väärtusi on ülimalt kahtlased, ei pruugi korraktsed olla ka ülejäänud ...

Peale andmetabeli muutmist on vajalik ümber arvutada ka kord juba leitud arvkarakteristikud. *Pivot Table*'i abil on see õnneks ülimalt lihtne:

PUDER 🔽	Data 🔽	Total					• klikite hiire p	arempoolse	
ei	Count of MASS	22					nunuga suval	isel Pivot Tak	ole'i
	Average of MASS	67,39090909	<u>о</u> р	Format Cells		1	lohtril io	150111101100	
	StdDev of MASS	21115228028	<u> </u>	_ormac cens.		_	lanun ja		
	Min of MASS2	1,8	III. F	Pivot <u>⊂</u> hart		-	• valite rippme	niiiist käsu	
	Max of MASS3	150	12 F	PivotTable <u>W</u>	izard	-	Refresh Data	Värskonda	andmaid
jah	MASS	35	9 F	Refresh Data	ı N	k i	Refresh Dulu	i vurskenuu i	unumetu.
	of MASS	15 69500101		-		\mathbf{N}			
- 1	MASS2	15,00550152		HIQE		$\langle \rangle$			
-	Max of MASS3	121		<u>5</u> elect	•				
nii ia naa	Count of MASS	125	9	Group and St	now Detail 🕨				
1	Average of MASS	78,8		Order			Λ		
	StdDev of MASS	6,833739825		oru <u>e</u> r		-	\backslash		
1	Min of MASS2	71	Ъ. Г	Field Setti <u>n</u> gs					
]	Max of MASS3	89	1	Table Option:	ş		· · · · · · · · · · · · · · · · · · ·		
Total Count of N	MASS	66		Hide PivotTał	PUDER	-	Data 💽	Total	_
Total Average of	f MASS	69,24393939		unde metal tra	ei		Count of MASS	21	
Total StdDev of	MASS	17,30499374		Hide Field Lisi	1		Average of MASS	70,52380952	
Total Min of MA	<u>.882</u>	1,6			1		StdDev of MASS	15,59044274	
LIOTAL MAX OF MA	4000	123					Min of MASS2	42	
							Max of MASS3	100	
					iah		Count of MASS	38	
					Jun		Average of MASS	67 64473684	
							StdDoy of MASS	13 11/31/30	
					-		Min of MACCO	10,11401400	
					-		Maria and MASSZ	40	
							INIAX OF MASS3	95	
					nii ja naa		Count of MASS	5	
							Average of MASS	78,8	
							StdDev of MASS	6,833739825	
							Min of MASS2	72	
]		Max of MASS3	89	
					Total Cou	nt of M	IASS	64	
					Total Ave	rage of	MASS	69,4609375	
					Total Stdl	Dev of I	MASS	13,79938597	
					Total Min	of MA	SS2	42	
					Total Max	of MA	SS3	100	

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	21	\backslash
		Average of MASS	70,52380952	_
		StdDev of MASS	15 59044274	-

	-				
	Data 🔽)			
PUDER 🔽	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	21	70,52380952	15,59044274	42	100
jah	38	67,64473684	13,11431439	46	95
nii ja naa	5	78,8	6,833739825	72	89
Grand Total	64	69,4609375	13,79938597	42	100



	Data 💽				
PUDER 💌	Count of MASS	Average of MASS	StdDev of MASS	Min of MASS2	Max of MASS3
ei	21	70,5	15,6	42	100
jah	38	67,6	13,1	46	95
nii ja naa	5	78,8	6,8	72	89
Grand Total	64	69,5	13,8	42	100

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		21	38	5	64
Average of MAS	S	70,5	67,6	78,8	69,5
StdDev of MASS	3	15,6	13,1	6,8	13,8
Min of MASS2		42	46	72	42
Max of MASS3		100	95	89	100

Aga järgnevale kujule?

				PUDER 🔽			
AINEKOOD	-	Data	-	ei	jah	nii ja naa	Grand Total
Hum	_	Count of MASS	_	8	19	2	29
]		Average of MAS	S	65,5	67,3	76,0	67,4
		StdDev of MASS	3	13,7	12,5	5,7	12,4
		Min of MASS2		44	46	72	44
		Max of MASS3		80	95	80	95
Muu		Count of MASS		1			1
		Average of MAS	S	62,0			62,0
		StdDev of MASS	3	#DIV/0!			#DIV/0!
]		Min of MASS2		62			62
		Max of MASS3		62			62
Reaal		Count of MASS		11	19	3	33
]		Average of MAS	S	74,5	68,0	80,7	71,3
		StdDev of MASS	3	17,5	14,1	8,0	15,1
]		Min of MASS2		42	48	73	42
		Max of MASS3		100	93	89	100
(blank)		Count of MASS		1			1
]		Average of MAS	S	75,0			75,0
]		StdDev of MASS	3	#DIV/0!			#DIV/0!
		Min of MASS2		75			75
		Max of MASS3		75			75
Total Count of N	1A3	SS		21	38	5	64
Total Average of	M	ASS		70,5	67,6	78,8	69,5
Total StdDev of	MA	4SS		15,6	13,1	6,8	13,8
Total Min of MA	SS	2		42	46	72	42
Total Max of MA	NS:	53		100	95	89	100

Et ebameeldivaim aine on nimetamata vaid ühel tudengil

(*Excel* tekitab puuduva väärtuse tarvis grupi '(blank)' e '(tühi)') ja ka ebameeldivaima aine koodile 'Muu' vastab andmestikus vaid üks kehamass, siis võiks vastavad ainekoodid tabelist üldse välja jätta:



			PUDER 🔽]		
AINEKOOD 💽	Data	-	ei	jah	nii ja naa	Grand Total
Hum	Count of MASS		8	19	2	29
	Average of MAS	S	65,5	67,3	76,0	67,4
	StdDev of MASS	S	13,7	12,5	5,7	12,4
	Min of MASS2		44	46	72	44
	Max of MASS3		80	95	80	95
Reaal	Count of MASS		11	19	3	33
	Average of MAS	S	74,5	68,0	80,7	71,3
	StdDev of MASS	3	17,5	14,1	8,0	15,1
	Min of MASS2		42	48	73	42
	Max of MASS3		100	93	89	100
Total Count of MASS		19	38	5	62	
Total Average of MASS		70,7	67,6	78,8	69,5	
Total StdDev of MASS		16,3	13,1	6,8	14,0	
Total Min of MASS2		42	46	72	42	
Total Max of MASS3		100	95	89	100	

10. Oskate te iga viimases tabelis sisalduva arvu kohta öelda, mida see näitab?

Proovige sõnastada paar järeldust keskmise kehamassi sarnasusest või erinevusest

- mannaputru söövatel ja mitte söövatel tudengitel ning
- humanitaar- ja reaalaineid ebameeldivaimaks pidanud tudengitel.