

Praktikum 1

MS Excelis on võimalik teostada suur hulk andmete haldamisest ja esmasest statistilisest analüüsist, sageli ka kogu vajalik analüüsides hulk.

Kuigi tänane praktikum käsitleb vaid erinevaid kirjeldava statistika võimalusi MS Excelis, võivad huvilised

- tutvuda ka põhjalikuma õppematerjaliga „Andmeandmeanalüüs MS Excel 2010 abil“ (http://www.eau.ee/~ktanel/andmeanalyys_excelis/),
- vaadata üle EMÜ töötajatele Exceli võimalusi tutvustava kursuse „Andmete haldus ja analüüs MS Excelis“ materjalid aadressil http://www.eau.ee/~ktanel/Exceli_koolitus_EMYS_2014/ või
- uurida EMÜ veterinaarmeditsiini ja loomakasvatuse instituudi esimese kursuse tudengite biomeetria-kursuse erinevatele teemadele pühendatud praktikume (http://www.eau.ee/~ktanel/VL_0781/praktikumid.php).

Tänases praktikumis analüüsitav andmestik, mis on osa Mariann Nõlvaku poolt aastail 2004-2006 kogutud Eesti kalade andmebaasist, sisaldab järgmisi tunnuseid:

- kala number (lihtsalt identifitseerimiseks);
- liik (6 liiki: haug, särg, latikas, luts, ahven ja koha);
- rühm: röövkala või lepiskala;
- 5 püügikohta (Võrtsjärv, Kärevere, Kastre, Praaga ja Peipsi järv);
- püügisesoon (kevad-suvi või sügis-talv);
- kaal ja pikkus;
- sugu;
- lõpuse-, silma- ja sooleparasiitide ning laiussi (*Diphyllbothrium latum*) leidude arv kalal (arv_lõpusepar, arv_soolepar, arv_silmapar ja arv_Diphyllob).
- lõpuse-, silma- ja sooleparasiitidega ning laiussiga nakatumine (kas jah või ei kodeerituna 1 ja 0);
- parasiitide arv kala kohta (par_per_kala).

Esimese asjana salvestagegi kursuse kodulehelt omale arvutisse andmestik *kalad.xlsx*. (http://www.eau.ee/~ktanel/DK_0007/kalad.xlsx).

ÜLDINE SISSEJUHATUS: PEAMISED ANDMEANALÜÜSI TEOSTAMISE VAHENDID MS EXCELIS

Järgmiste näidete aluseks olev 12 lehma andmeid sisaldav Exceli fail on allalaaditav aadressilt

<http://www.eau.ee/~ktanel/>

[Exceli koolitus EMYs 2014/lehmad 1.xlsx](#)

See peatükk annab vaid kiirülevaate, et ühtteist meelde tuletada või anda mõningaid esmaseid näpunäiteid. Pisut täpsemalt tuleb kirjeldavast statistikast *Excelis* juttu edaspidi ja ülejäänud *Exceli* võimalustega saate tutvuda eelmisel lehel viidatud materjalide abil.

	A	B	C	D	E	F
1	Lehm	Farm	Sünniaeg	Tõug	Piim, kg	Rasv, %
2	Roosi	1023	22.08.2006	EHF	55,1	5,00
3	Mari	1023	20.08.2007	EHF	37,7	2,87
4	Mimmi	1023	22.03.2007	EHF	24,4	4,75
5	Mustik	1023	14.06.2007	EHF	30,9	4,10
6	Killu	1056	17.10.2006	EHF	40,7	4,34
7	Liisu	1023	26.06.2007	EPK	24,6	4,91
8	Tibi	1023	29.10.2006	EPK	30,1	1,83
9	Tooru	1023	28.05.2008	EPK	19,3	2,92
10	Saepuruu	1056	7.01.2008	EK	26,4	4,16
11	Kirka	1056	29.08.2006	EK	22,9	4,90
12	Labiidas	1056	2.03.2008	EK	15,5	4,04
13	Stina	1056	13.09.2007	EPK	34,6	4,30

Nipid, märkused, soovitused.

Andmetabeli ülesehitusest ja arvutustest.

- Rangelt soovitatav on (st, et see on ilmtingimata vajalik mõnede analüüside teostamiseks ja teeb teiste puhul andmete ette andmise lihtsamaks)
 - paigutada andmetabel töölehe ülemisse vasakusse nurka ning
 - kirjutada andmetabeli esimesse (ja üksnes esimesse) ritta iga veeru tarvis tunnuste nimed, mis identifitseerivad veerud üheselt.
- Mistahes arvutustulemuste-kokkuvõtete-kommentaare ning algse andmetabeli vahele peaks jääma vähemalt üks tühi rida ja veerg, sest vastasel juhul käsitleb Excel mitmete operatsioonide puhul neid kommentaare jmt andmetabeli osadena, mis omakorda võib viia valede arvutustulemusteni.

Järgnevalt on esitatud vaikumisi analüüstav andmetabel sorteerimisel, filtreerimisel ja *PivotTable*'i rakendamisel, kui operatsiooni valimise hetkel paiknes kursor lahtris A1:

	A	B	C	D	E
1	Loom	Sugu	Pikkus		
2	1	e	54,3	Memo!	
3	2	e	34,2	Mine poodi ja too saia	
4	3	i	31,1		
5	4	i	28,7		
6		Keskmine	37,075		
7					


	A	B	C	D	E	F
1	Loom	Sugu	Pikkus			
2	1	e	54,3		Memo!	
3	2	e	34,2		Mine poodi ja too saia	
4	3	i	31,1			
5	4	i	28,7			
6						
7		Keskmine	37,075			

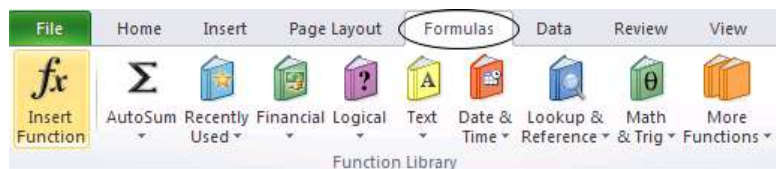
--- Funktsioonid ---

Samm 1.

Pange kursor sellesse lahtrisse, kuhu funktsiooni tulemust soovite. Näiteks soovides arvutada lehmade keskmist piimatoodangut lahtrisse E15, tuleb sinna panna ka kursor.

Samm 2.

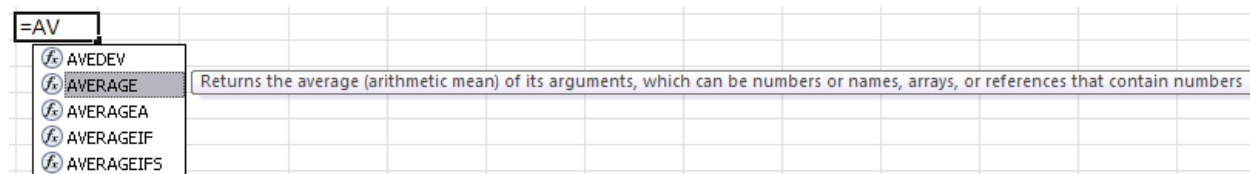
- Klikkige valemi rea ees oleval nupul  ning valige avanenud aknas **funktsiooni tüüp ja nimi**
- või valige soovitud funktsioon *Formulas*-sakilt:



- või sisestage funktsioon klaviatuurilt ise, trükkides esmalt võrdusmärgi = ja selle järel funktsiooni nime.

See viimane versioon on kiireim ja mugavam!

Isegi kui te ei mäleta funktsiooni täpset nime ja süntaksit, piisab, kui trükkite võrdusmärgi ja selle järel funktsiooni nime esimese tähe – Excel pakub teile valikut kõigist antud tähega algavaist funktsioonidest, lisades järgmise tähe, funktsioonide valik kitseneb. Lisaks kuvab Excel selle funktsiooni kohta, millel hiirega klikite, ka lühiselgituse:

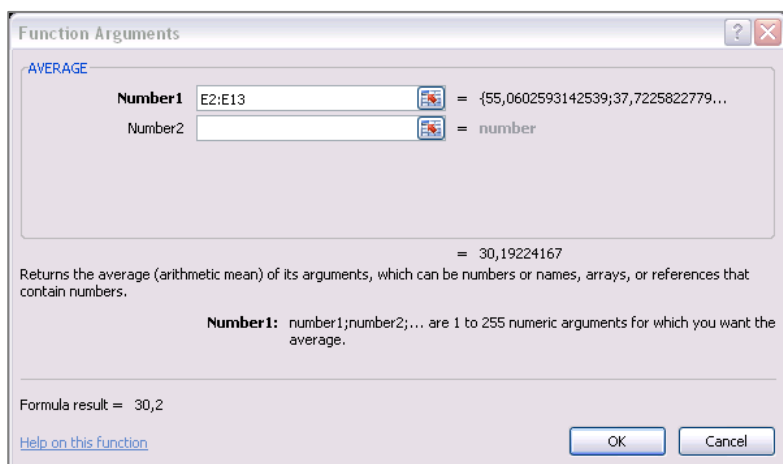


Samm 3.

Andke ette funktsiooni argumendid, klikkides vastavatel lahtritel või trükkides lahtrite aadressid või vastavad väärtused klaviatuurilt.

NB! Kui te hakkasite funktsiooni sisestama seda ise trükkides, tuleb funktsiooni valimiseks Exceli poolt kuvatud nimekirjast teha sellel hiirega topeltklakk.

Tulemus:



=AVERAGE(

AVERAGE(number1; [number2]; ...)

A...	A	B	C	D	E	F
1	Lehm	Farm	Sünniaeg	Tõug	Piim, kg	Rasv, %
2	Roosi	1023	22.08.2006	EHF	55,1	5,00
3	Mari	1023	20.08.2007	EHF	37,7	2,87
4	Mimmi	1023	22.03.2007	EHF	24,4	4,75
5	Mustik	1023	14.06.2007	EHF	30,9	4,10
6	Killu	1056	17.10.2006	EHF	40,7	4,34
7	Liisu	1023	26.06.2007	EPK	24,6	4,91
8	Tibi	1023	29.10.2006	EPK	30,1	1,83
9	Tooru	1023	28.05.2008	EPK	19,3	2,92
10	Saepuruu	1056	7.01.2008	EK	26,4	4,16
11	Kirka	1056	29.08.2006	EK	22,9	4,90
12	Labiidas	1056	2.03.2008	EK	15,5	4,04
13	Stina	1056	13.09.2007	EPK	34,6	4,30
14						
15					=AVERAGE(E2:E13)	

Nipid, märkused, soovitused.

- Kui lahtri sisu algab võrdusmärgiga, ootab Excel selle järel alati käsku, mida teostada. Näiteks trükkides lahtrisse H2 käsu =3*2 ja vajutades *Enter*-klahvi, on tulemuseks väärtus 6. Soovides lahtris nähagi teksti kujul =3*2, tuleb võrdusmärgi ette trükkida ülakoma, st. et lahtri sisu peab olema '=3*2'.
- Arvutuste aluseks olevaid väärtuseid sisaldavaid lahtreid võite funktsioonidele ette anda
 - vastavaid lahtreid töölehel hiirega valides;
 - klikkides lahtribloki viimasel lahtril ja seejärel Shift-klahvi all hoides lahtribloki esimesel lahtril – **see on sageli kiireim viis suurte andmetabelite puhul**, sest kerimisriba abil saate kiiresti liikuda tabeli algusesse;
 - trükkides klaviatuurilt analüüsitavate lahtrite aadressid; seejuures tähendab koolon kahe lahtri aadressi vahel 'kuni' (E2:E13 tähendab lahtreid E2 kuni E13), semikoolon kahe lahtri aadressi vahel tähendab aga 'ja' (E2;E13 tähendab lahtreid E2 ja E13), argumentide esitus kujul E7:E9;E13 tähendab aga lahtreid E7 kuni E9 pluss E13 – näiteks arvutab valem =AVERAGE(E7:E9;E13) eesti punast tõugu lehmade keskmise piimatoodangu algsest lehmade andmetabelist.
- Igasugu kokkuvõtlike tulemuste ja algse andmetabeli vahele peaks alati jätma vähemalt ühe tühja rea ja veeru – vajalik on see selleks, et Excel ei tõlgendaks kokkuvõtlikke suuruseid andmetabeli osana (näiteks ühe täiendava lehma piimatoodanguna).
- **NB! Semikoolonit** kasutatakse erinevate lahtrite valimiseks (näiteks E2;E13) ja üldisemalt, funktsiooni erinevate argumentide eraldamiseks (näiteks =COUNTIF(F2:F13;">4")), vaid **eesti keele seadistuses** Exceli puhul, **inglise keele seadistuses** Excelis on funktsiooni argumentide eraldajaks **koma!**

--- Risttabel (Pivot Table) ---**Samm 1.**

Pange kursor suvalisse andmetabeli lahtrisse – eeldustel, et

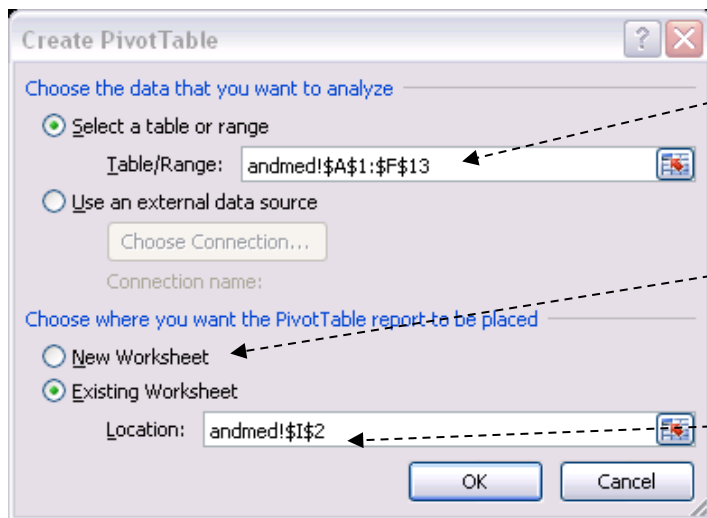
- **andmetabeli esimeses (ja üksnes esimeses!) reas on tunnuste nimed ning**
- **analüüsitav andmetabel on muudest arvutus-tulemustest ja kommentaaridest eraldatud vähemalt ühe tühja rea ja veeruga,**

kaasab Excel analüüsi kogu andmetabeli.

	A	B	C	D	E	F
1	Lehm	Farm	Sünniaeg	Tõug	Piim, kg	Rasv, %
2	Roosi	1023	22.08.2006	EHF	55,1	5,00
3	Mari	1023	20.08.2007	EHF	37,7	2,87
4	Mimmi	1023	22.03.2007	EHF	24,4	4,75
5	Mustik	1023	14.06.2007	EHF	30,9	4,10
6	Killu	1056	17.10.2006	EHF	40,7	4,34
7	Liisu	1023	26.06.2007	EPK	24,6	4,91
8	Tibi	1023	29.10.2006	EPK	30,1	1,83
9	Tooru	1023	28.05.2008	EPK	19,3	2,92
10	Saepuruu	1056	7.01.2008	EK	26,4	4,16
11	Kirka	1056	29.08.2006	EK	22,9	4,90
12	Labiidas	1056	2.03.2008	EK	15,5	4,04
13	Stina	1056	13.09.2007	EPK	34,6	4,30

Samm 2.

Insert-sakk → PivotTable



- Kui te unustasite algselt kursori andmetabelisse panemata või soovite lihtsalt täpsustada analüüstavat andmetabelit, saate seda teha siin.
- Vaikimisi lisatakse *PivotTable*'i tarvis *Exceli* tööraamatusse uus tööleht (*New Worksheet*) ning paigutatakse konstrueeritav tabel sinna.
- *PivotTable*'i mõnele juba olemas olevale lehele paigutamiseks tuleb ära märkida teine valik (*Existing Worksheet*) ning anda ette loodava tabeli vasaku

ülemise nurga aadress (kas klikkides hiirega vastaval tühjal lahtril või trükkides lahtri aadressi kasti *Location*).

Samm 3.Avanened *PivotTable*'i konstrueerimise aknas lohistage

- kasti *Row Labels* tunnus(ed), mille väärtuste alusel soovite jagada tabelit ridadeks,
- kasti *Column Labels* tunnus(ed), mille väärtuste alusel soovite jagada tabelit veergudeks,
- kasti Σ *Values* tunnus(ed), mille väärtuste kohta soovite midagi arvutada ning
- kasti *Report Filter* tunnus(ed), mille väärtuste alusel soovite konstrueerida erinevaid tabelleid.

NB! Kõik nimetatud lahtrid ei pea olema täidetud.

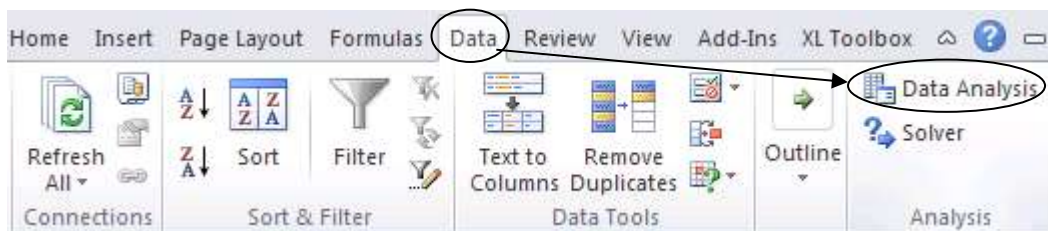
Näiteks soovides lugeda kokku, kui mitu mingit tõugu lehma mingist farmist on, võib tabeli konstrueerida järgmiselt:

	Farm	EHF	EK	EPK	Grand Total
Count of Tõug	1023	4	3	3	7
Tõug	1056	1	3	1	5
Grand Total		5	3	4	12

--- Statistikaprotseduurid ---

Samm 1.

Data-sakk → Data Analysis

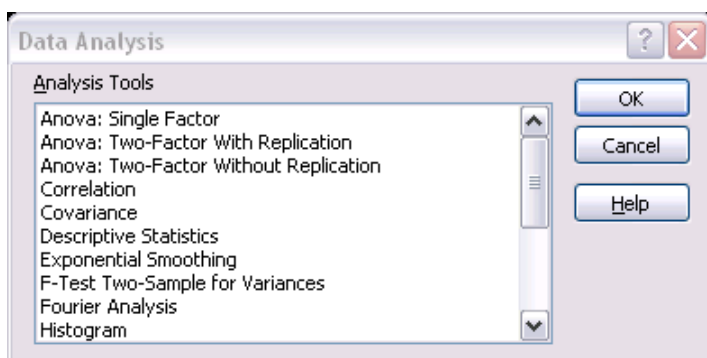


NB! Kui valik *Data Analysis* teie arvutis puudub, tuleb see aktiveerida (välja kutsuda):

File → *Options* → *Add-Ins* → *Manage [Excel Add-ins] [Go...]* → *Analysis ToolPak*.

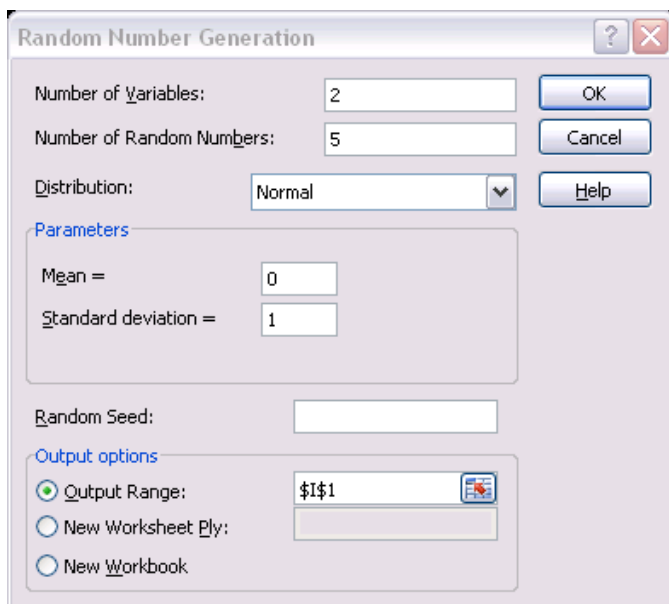
Samm 2.

Avanenud aknas valige vajalikku analüüsi teostav protseduur ning



- ande sellele ette analüüsitavaid andmeid sisaldavate lahtrite aadressid (mitte kõigil protseduuridel),
- määrake väljuntabelite-jooniste asukoht (vaikimisi uuele töölehele) ning
- täpsustage teostatavad analüüsid ja väljastatavad tabelid (osadel protseduuridel).

Näiteks soovides genereerida kahte veergu kummassegi viit standardse normaaljaotusega juhuslikku arvu ning paigutada tulemused avatud töölehele alates lahtrist I1, tuleb statistikaprotseduuride hulgast valida *Random Number Generation*, täita ära vastavad lahtrid ja klikkida nupul OK:



Tulemus:

I	J
-0,58056	0,083057
-0,52977	-1,073
0,205966	0,50068
0,119276	-0,01549
-0,2572	-1,28726

Nipid, märkused, soovitused.

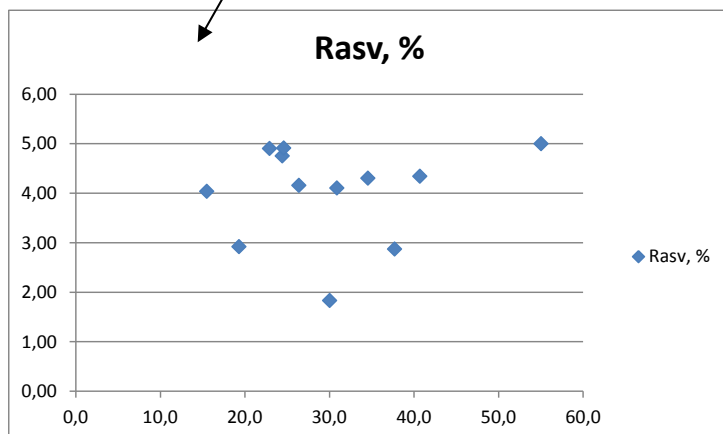
- Kui funktsioonid annavad enamasti tulemuseks ühe väärtuse (va üksikud massiivi-funktsioonid), on statistikaprotseduuride tulemuseks mitmed mahukad tabelid ja mõnikord ka joonised.
- Kursori paiknemisest analüüside teostamise eel:
 - funktsioonide rakendamisel peab kursor olema (tühjas) lahtris, kuhu soovite tulemust saada,
 - PivotTable'i konstrueerimisel on soovitatav panna kursor andmetabeli sisse,
 - statistikaprotseduuride rakendamisel ei ole kursori asukoht tähtis (andmeid on mugavam ette anda, kui aktiivne on analüüsitavaid andmeid sisaldav tööleht).
- Tunnuste nimede arvestamisest:
 - funktsioonid eeldavad reeglina, et andmed on ette antud ilma tunnuse nimeta,
 - *PivotTable* eeldab vastupidi, et andmetabeli esimeses reas on (unikaalsed) tunnuste nimed,
 - statistikaprotseduurid, mis opereerivad Exceli töölehel kirjas olevate väärtustega, tunnuste nimede olemasolu andmetabeli esimeses reas ei eelda, aga võimaldavad nende olamsolul andmeid analüüsida koos nimedega (siis tuleb ära märkida lisavalik *Labels in first row*) ning kasutavad neid tulemuste väljatrükis.

--- Joonised ---

Samm 1.

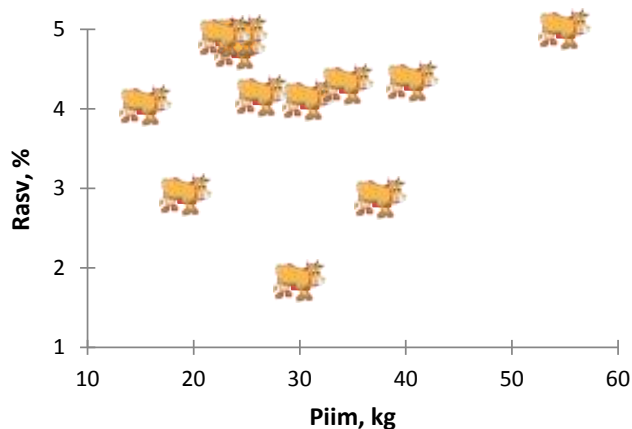
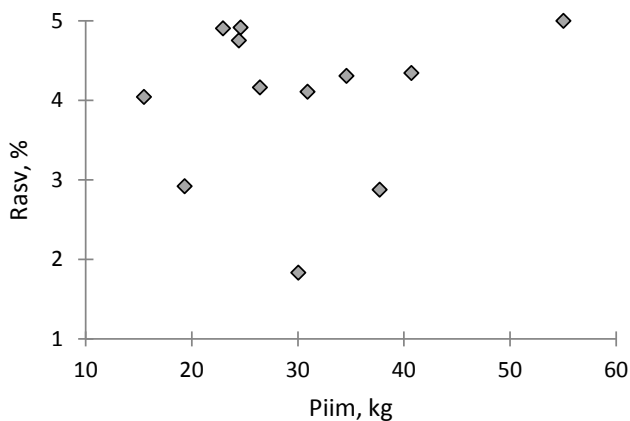
Võtke blokki andmed, mille alusel soovite joonist konstrueerida ning valige *Insert*-sakilt sobiv joonis. Näiteks:

	A	B	C	D	E	F
1	Lehm	Farm	Sünniaeg	Tõug	Piim, kg	Rasv, %
2	Roosi	1023	22.08.2006	EHF	55,1	5,00
3	Mari	1023	20.08.2007	EHF	37,7	2,87
4	Mimmi	1023	22.03.2007	EHF	24,4	4,75
5	Mustik	1023	14.06.2007	EHF	30,9	4,10
6	Killu	1056	17.10.2006	EHF	40,7	4,34
7	Liisu	1023	26.06.2007	EPK	24,6	4,91
8	Tibi	1023	29.10.2006	EPK	30,1	1,83
9	Tooru	1023	28.05.2008	EPK	19,3	2,92
10	Saepuruu	1056	7.01.2008	EK	26,4	4,16
11	Kirka	1056	29.08.2006	EK	22,9	4,90
12	Labiidas	1056	2.03.2008	EK	15,5	4,04
13	Stina	1056	13.09.2007	EPK	34,6	4,30



Samm 2.

Kujundage joonis sobivaks.



OSA 1

--- Arvkarakteristikud ---

❖ Funktsioonide kasutamine

1) Leidke andmetabeli alla püütud ahvenate arv, keskmine pikkus, pikkuse standardhälve ja standardviga, mediaan, alumine ja ülemine kvartiil, minimaalne ja maksimaalne väärtus.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
kala nr	koht	sesoon	sugu	kaal	pikkus	arv	kala nr	liik	ryhm	pyygi	koht	sesoon	sugu	kaal	pikkus	arv	lõpuse arv	silms
1010	7				144	24	2	374	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	131	22,5	10	2	0
1011	7				108	22	4	372	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	202	25	0	4	0
1012	6				135	23	5	560	Ahven	rookkala	Peipsi	sygis-talv	e	134	21,5	6	4	0
1013	6				150	23,5	6	377	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	84	18,5	6	6	0
1014	6				123	22,5	7	365	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	145	23	6	3	0
1015	6				92	20	8	335	Ahven	rookkala	Praaga	kevad-suvi	i	85	19,5	20	1	0
1016	6				103	20,5	9	64	Ahven	rookkala	Kastre	kevad-suvi	e	166	23	24	1	0
1017	6				109	21	10	527	Ahven	rookkala	Võrtsjärv	sygis-talv	i	152	22	2	2	0
1018	6				117	22	11	525	Ahven	rookkala	Võrtsjärv	sygis-talv	e	218	24,5	2	4	0
1019	6				117	22	12	376	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	111	22	4	3	0
1020	6				124	22,5	13	375	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	132	23	0	0	0
1021	6				117	21,5	14	360	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	351	31	22	0	0
1022	6				146	22,5	15	347	Ahven	rookkala	Praaga	kevad-suvi	i	70	19	14	0	0
1023	5				129	22	16	344	Ahven	rookkala	Praaga	kevad-suvi	e	61	18	30	1	0
1024	5				126	22	17	343	Ahven	rookkala	Praaga	kevad-suvi	i	93	20	14	0	0
1026	5				156	22,5	18	310	Ahven	rookkala	Peipsi	kevad-suvi	e	87	19	18	1	0
1027	5				133	21,5	19	755	Ahven	rookkala	Peipsi	kevad-suvi	i	133	22	24	3	0
1028					267	28	20	754	Ahven	rookkala	Peipsi	kevad-suvi	i	194	23,5	0	3	0
1029					116	22	21	753	Ahven	rookkala	Peipsi	kevad-suvi	i	192	23,5	2	0	0
1030	10				121	21	22	521	Ahven	rookkala	Võrtsjärv	sygis-talv	i	229	24	0	3	0
1031	10				76	18	23	487	Ahven	rookkala	Kastre	sygis-talv	e	45	15	2	1	0
1032	10				89	19,5	24	373	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	141	21	0	3	0
1033	10				138	22	25	371	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	248	27	4	11	0
1034	10				147	22	26	361	Ahven	rookkala	Võrtsjärv	kevad-suvi	e	250	27,5	10	1	0
1035	10				137	22,5	27	342	Ahven	rookkala	Praaga	kevad-suvi	i	43	15,5	18	6	0
1036	10				126	22	28	341	Ahven	rookkala	Praaga	kevad-suvi	i	58	18	26	0	0
1037	9				33	13	29	339	Ahven	rookkala	Praaga	kevad-suvi	i	70	18,5	28	3	0
1038	9				32	13	30	334	Ahven	rookkala	Praaga	kevad-suvi	e	123	22	28	0	0
1039	904	Särg	lepiskala	Peipsi	sygis-talv	e	31	329	Ahven	rookkala	Kastre	kevad-suvi	e	76	18	50	0	0
1040	953	Särg	lepiskala	Peipsi	sygis-talv	i	32	319	Ahven	rookkala	Peipsi	kevad-suvi	e	80	18,5	72	0	0
1041	933	Särg	lepiskala	Praaga	sygis-talv	e	33	318	Ahven	rookkala	Peipsi	kevad-suvi	e	77	18,5	4	0	0
1042	651	Särg	lepiskala	Kastre	kevad-suvi	i	34	316	Ahven	rookkala	Peipsi	kevad-suvi	e	61	18	7	0	0
1043	636	Särg	lepiskala	Kastre	kevad-suvi	e	35											

1046	Ahven	kalade arv	=COUNT(H2:H180)	179
1047		keskmine	=AVERAGE(H2:H180)	20,45
1048		standardhälve	=STDEV(H2:H180)	3,55
1049		standardviga	=H1048/SQRT(H1046)	0,27
1050		mediaan	=MEDIAN(H2:H180)	20
1051		alumine kvartiil	=QUARTILE.INC(H2:H180;1)	18
1052		ülemine kvartiil	=QUARTILE.INC(H2:H180;3)	23
1053		min	=MIN(H2:H180)	14
1054		max	=MAX(H2:H180)	32

Nipid, märkused, soovitused.

Excelis puudub funktsioon aritmeetilise keskmise standardvea arvutamiseks. Aga teades standardvea arvutamise valemit: $se = s/\sqrt{n}$ (keskmise standardviga võrdub standardhälve jagatud ruutjuurega vaatluste arvust), saab selle vajadusel ikkagi arvutada (st ülalolev joonis).

2) Püüdke sõnastada leitud suuruste alusel vähemalt kaks lauset (kasutades vähemalt ühes mediaani ja/või kvartiile).

3) Leidke samad suurused ka ahvenate massile.

Ahven	kalade arv	179
	keskmine	20,45
	standardhälve	3,55
	standardviga	0,27
	mediaan	20
	alumine kvartiil	18
	ülemine kvartiil	23
	min	14
	max	32

- 4) Arvutage laiussi noorvormide arv keskmiselt kõigi ahvenate kohta ja vaid nakatunud ahvenate kohta. Kui mitu % ahvenatest üleüldse olid laiussiga nakatunud?

Järgnevalt esitatud spikrit kasutades püüdke aru saada ka kasutatavate funktsioonide loogikast ... (kui aru ei saa, küsi!)

Formula bar: `=SUM(L2:L180)/COUNTIF(L2:L180;">0")`

	K	L	M	N	O	P	Q	R	S	T	
1	arv	soolep	arv	Diphy	l6pusepar	silmapar	soolepar	Diphylob	par	per	kala
1040		0	0	0	1	0	0	0	1		
1041		0	0	0	1	0	0	0	1		
1042		0	0	0	1	0	0	0	1		
1043		0	0	0	1	0	0	0	1		
1044											
1045		Keskmine laiussi noorvormide arv nakatunutel:					Laiussi noorvormidega nakatunud kalade osakaal				
1046		=SUM(L2:L180)/COUNTIF(L2:L180;">0")					(alternatiivne variant leidmiseks):				
1047		Keskmine laiussi noorvormide arv:					=AVERAGE(P2:P180)				
1048		=AVERAGE(L2:L180)									
1049		Laiussi noorvormidega nakatunud kalade osakaal:									
1050		=COUNTIF(L2:L180;">0")/COUNT(L2:L180)									

↓

Keskmine laiussi noorvormide arv nakatunutel:	1,3333	Laiussi noorvormidega nakatunud kalade osakaal
		(alternatiivne variant leidmiseks):
Keskmine laiussi noorvormide arv:	0,0223	0,01676
Laiussi noorvormidega nakatunud kalade osakaal:	0,01676	

↓

Keskmine laiussi noorvormide arv nakatunutel:	1,3333	Laiussi noorvormidega nakatunud kalade osakaal
		(alternatiivne variant leidmiseks):
Keskmine laiussi noorvormide arv:	0,0223	1,68%
Laiussi noorvormidega nakatunud kalade osakaal:	1,68%	

Number format dialog: Number, %, .00, .00, .0

- 5) Ja nüüd arvutage samad suurused ka haugide tarvis (kasutades maksimaalselt ära juba rakendatud funktsioone ja vorminguid – see tähendab, et tehke koopia kõigest ahvenate kohta kirjutatust ja arvutatust, edasi jääb üle muuta vaid funktsioonide argumente).

Ahven					Keskmine laiussi noorvormide arv nakatunutel:	Laiussi noorvormidega nakatunud kalade osakaal:
kalade arv		179	179		1,3333	1,68%
keskmine		118,60	20,45		Keskmine laiussi noorvormide arv:	
standardhälve		73,00	3,55		0,0223	
standardviga		5,46	0,27		Laiussi noorvormidega nakatunud kalade osakaal:	
mediaan		95	20		1,68%	
alumine kvartiil		70	18			
ülemine kvartiil		151,5	23			
min		34	14			
max		517	32			
Haug					Keskmine laiussi noorvormide arv nakatunutel:	Laiussi noorvormidega nakatunud kalade osakaal:
kalade arv		214	214		1,3214	13,08%
keskmine		785,63	48,76		Keskmine laiussi noorvormide arv:	
standardhälve		339,53	5,84		0,1729	
standardviga		23,21	0,40		Laiussi noorvormidega nakatunud kalade osakaal:	
mediaan		738	48,25		13,08%	
alumine kvartiil		578,75	45			
ülemine kvartiil		881	51,5			
min		238	26			
max		3013	74			

❖ Olulisemate arvarakteristikute leidmise teisi variante

Kuigi *MS Exceli* funktsioonidel on hulk positiivseid omadusi:

- argumentide ette andmine on enamasti intuitiivselt mõistetav;
- tänu võimalusele funktsioone kopeerida on kord juba sisestatud käsud lihtsalt rakendatavad uutele argumentidele (väärtustele, tunnustele);
- funktsioonide omavaheline kombineerimine võimaldab väljastada keeruliste avaldiste tulemusi;

on teatud olukordades siiski sobivam kasutada mõnda teist *MS Exceli* vahendit.

Näiteks

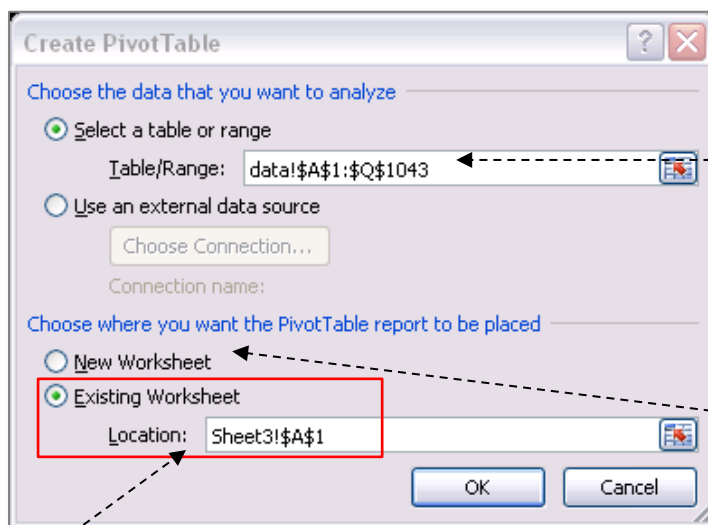
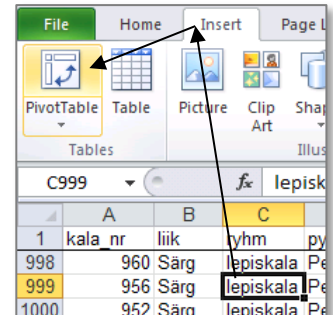
1. arvutamaks keskmisi väärtusi gruppides (näiteks keskmisi pikkusi erinevat liiki kaladel) on mugav kasutada *PivotTable*'t (sakilt *Insert*);
2. leidmaks ühe korruga suurt hulka arvarakteristikuid, samuti usalduspiire keskmisele, võib kasutada statistikaprotseduuri *Descriptive Statistics (Data-sakk -> Data Analysis ->)*;
3. saamaks kiirelt teada mõne arvarakteristiku väärtust, ilma seda kuhugi töölehe lahtrisse arvutamata, võib kasutada töölehe allservas (olekuribal) kuvatavaid selekteeritud lahtrite sisu kirjeldavaid väärtusi.

▪ *PivotTable*

- 1) Arvutage kõigi kalaliikide kohta püütud kalade arv, keskmine, minimaalne ja maksimaalne pikkus ning laiussi noorvormiga nakatunud kalade osakaal vahendi *PivotTable* abil.

--- Tööjuhend ---

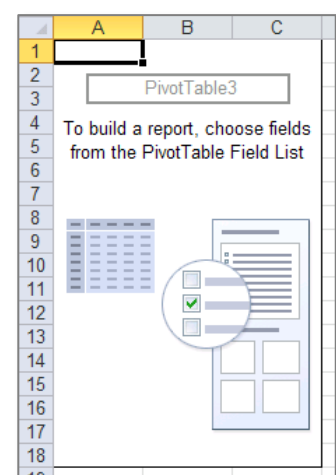
- Paigutage kursor andmetabeli suvalisse lahtrisse.
- *Insert*-sakk → *PivotTable*
- Loodav tabel paigutage töölehe 'Sheet3' ülemisse vasakusse nurka (hiljem nimetage tööleht 'Sheet3' ümber leheks 'PivotTable').



α) Kui paigutasite kursori enne käsu *PivotTable* valimist andmetabeli suvalisse lahtrisse, võttis *Excel* tabeli konstrueerimisel vaikimisi aluseks teie andmetabeli ja midagi enesel määrata pole vaja.

α) Vaikimisi lisatakse *PivotTable*'i tarvis *Exceli* tööraamatusse uus tööleht (*New Worksheet*) ning paigutatakse konstrueeritav tabel sinna.

PivotTable'i mõnele juba olemas olevale lehele paigutamiseks tuleb ära märkida teine valik (*Existing Worksheet*) ning anda ette loodava tabeli vasaku ülemise nurga aadress (kas klikkides hiirega vastaval tühjal lahtril või trükkides lahtri aadressi kasti *Location*).



○

To build a report, choose fields from the PivotTable Field List

Lohistage tunnus 'liik' lahtrisse Row Labels ning tunnus 'kala_nr' lahtrisse Σ Values

Tulemus:

Row Labels	Sum of kala_nr
Ahven	61256
Haug	69687
Koha	43300
Latikas	139832
Luts	62075
Särg	183458
Grand Total	559608

- *Exceli* poolt vaikumisi leitud kalade numbrite summa asemel lihtsalt väärtuste arvu tellimiseks on mitmeid erinevaid võimalusi. Näiteks:

Value Field Settings

Source Name: kala_nr

Custom Name: Count of kala_nr

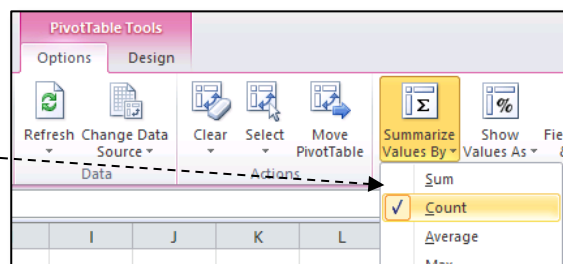
Summarize Values By: Show Values As

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

Sum
Count
Average
Max
Min
Product

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

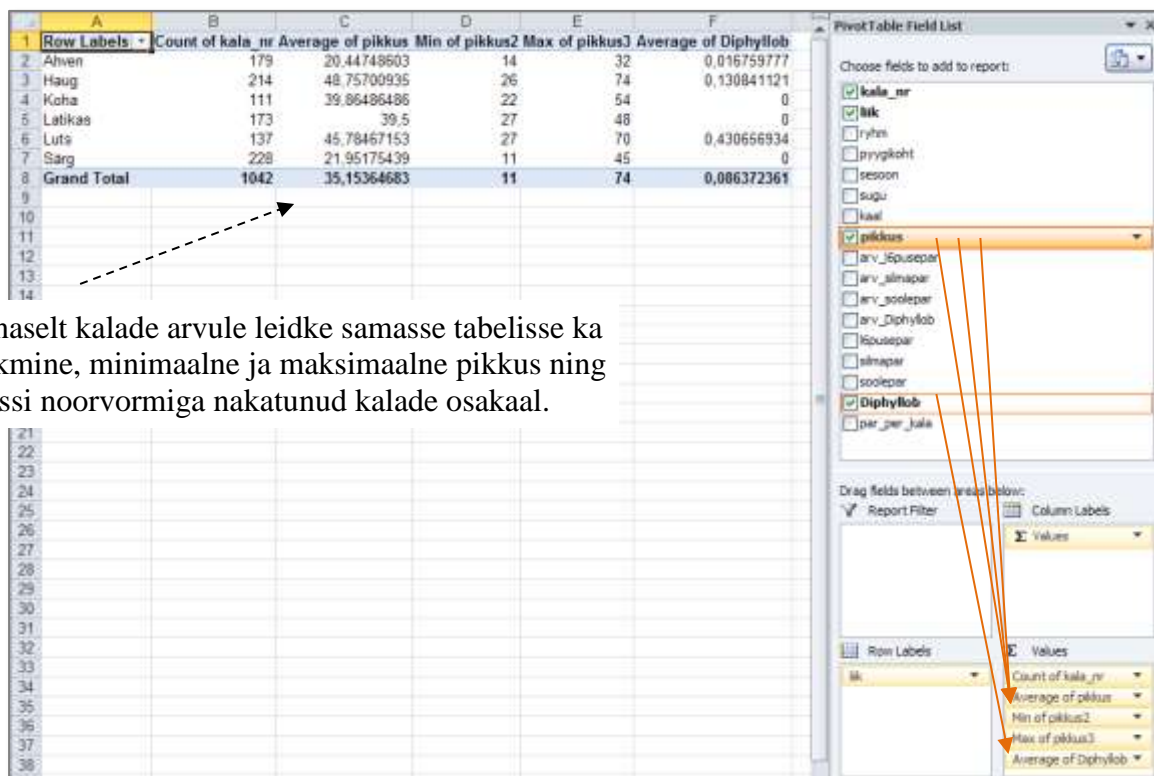
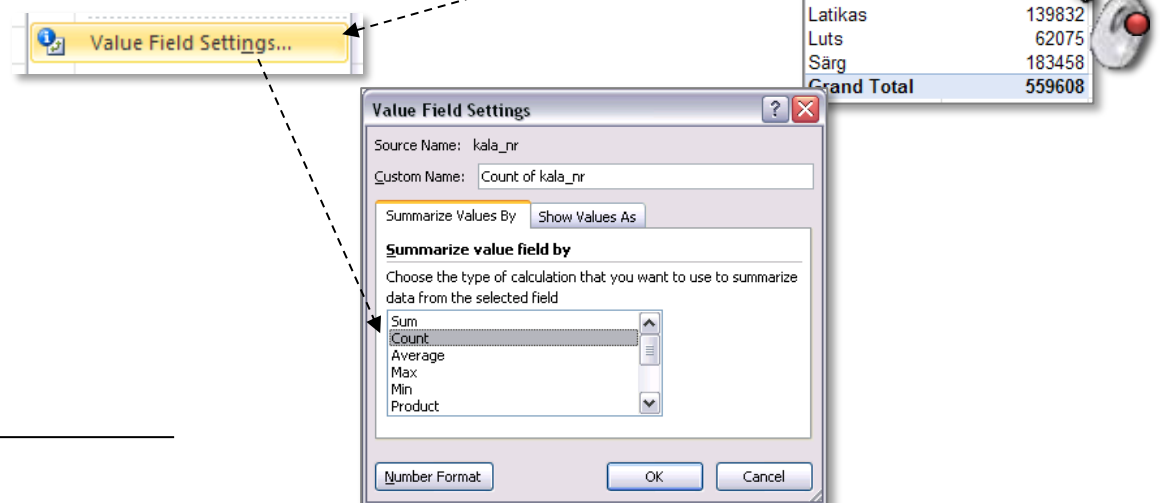


☞ Rakendatavat funktsiooni saab muuta ka klikkides muudetaval tunnusel

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:



- Sarnaselt kalade arvule leidke samasse tabelisse ka keskmine, minimaalne ja maksimaalne pikkus ning laiussi noorvormiga nakatunud kalade osakaal.

2) Ümardage keskmised ühe kümnendkohani ja esitage laiussi noorvormide esinemissagedused protsentides (kahe kümnendkohaga).

3) Edasi püüdke tabeliga pisut mängida, et tutvuda/meenutada mõningaid PivotTable võimalusi.
Näiteks proovige, kas oskate tabelit esitada järgmistel erinevatel kujudel.

Row Labels	Count of kala_nr	Average of pikkus	Min of pikkus2	Max of pikkus3	Average of Diphyllob
Ahven	179	20,4	14	32	1,68%
Haug	214	48,8	26	74	13,08%
Koha	111	39,9	22	54	0,00%
Latikas	173	39,5	27	48	0,00%
Luts	137	45,8	27	70	43,07%
Särg	228	22,0	11	45	0,00%
Grand Total	1042	35,2	11	74	8,64%

Values	Ahven	Haug	Koha	Latikas	Luts	Särg	Grand Total
Count of kala_nr	179	214	111	173	137	228	1042
Average of pikkus	20,4	48,8	39,9	39,5	45,8	22,0	35,2
Min of pikkus2	14	26	22	27	27	11	11
Max of pikkus3	32	74	54	48	70	45	74
Average of Diphyllob	1,68%	13,08%	0,00%	0,00%	43,07%	0,00%	8,64%

Row Labels	Count of kala_nr	Average of pikkus	Min of pikkus2	Max of pikkus3	Average of Diphyllob
lepiskala	401	29,5	11	48	0,00%
Latikas	173	39,5	27	48	0,00%
Särg	228	22,0	11	45	0,00%
roovkala	641	38,7	14	74	14,04%
Ahven	179	20,4	14	32	1,68%
Haug	214	48,8	26	74	13,08%
Koha	111	39,9	22	54	0,00%
Luts	137	45,8	27	70	43,07%
Grand Total	1042	35,2	11	74	8,64%

Row Labels	Count of kala_nr	Average of pikkus	Min of pikkus2	Max of pikkus3	Average of Diphyllob
Ahven	179	20,4	14	32	1,68%
Haug	214	48,8	26	74	13,08%
Koha	111	39,9	22	54	0,00%
Latikas	173	39,5	27	48	0,00%
Luts	137	45,8	27	70	43,07%
Särg	228	22,0	11	45	0,00%
Total Count of kala_nr	1042				
Total Average of pikkus		35,2			
Total Min of pikkus2			11		
Total Max of pikkus3				74	
Total Average of Diphyllob					8,64%

4) Sorteerige tabel kalaliikide keskmiste pikkuste järgi rühmade (lepiskala ja roovkala) siseselt.

Row Labels	Count of kala_nr	Average of pikkus	Min of pikkus2	Max of pikkus3	Average of Diphyllob
lepiskala	401	29,5	11	48	0,00%
Latikas	173	39,5	27	48	0,00%
Särg	228	22,0	11	45	0,00%
roovkala	641	38,7	14	74	14,04%
Haug	214	48,8	26	74	13,08%
Luts	137	45,8	27	70	43,07%
Koha	111	39,9	22	54	0,00%
Ahven	179	20,4	14	32	1,68%
Grand Total	1042	35,2	11	74	8,64%

5) Jätke tabelist välja kalaliigid, kelle puhul ei täheldatud laiussiga nakatumist (so lepiskalad ja roovkaladest koha).

Row Labels	Count of kala_nr	Average of pikkus	Min of pikkus2	Max of pikkus3	Average of Diphyllob
roovkala	530	38,4	14	74	16,98%
Haug	214	48,8	26	74	13,08%
Luts	137	45,8	27	70	43,07%
Ahven	179	20,4	14	32	1,68%
Grand Total	530	38,4	14	74	16,98%

- **Protseduur *Descriptive Statistics***

1) Arvutage nii palju erinevaid arvarakteristikuid ahvenate pikkuse ja massi kohta, kui protseduur *Descriptive Statistics* võimaldab.

Võimalik on analüüsida mitut tunnust korraga tingimusel, et nende väärtused paiknevad kõrvuti veergudes

← Kas siia kastikesse on „linnukest“ vaja?

Väljundtabeli vasaku ülemise nurga asukoht

Lisavalikute 'Summary statistics' jt kohta vt järgmine lk.

<i>kaal</i>		<i>pikkus</i>	
Mean	118,6	Mean	20,44749
Standard Error	5,4561	Standard Error	0,26514
Median	95	Median	20
Mode	75	Mode	18
Standard Deviation	72,997	Standard Deviation	3,547326
Sample Variance	5328,6	Sample Variance	12,58352
Kurtosis	5,9845	Kurtosis	0,328945
Skewness	2,034	Skewness	0,771968
Range	483	Range	18
Minimum	34	Minimum	14
Maximum	517	Maximum	32
Sum	21230	Sum	3660,1
Count	179	Count	179
Largest(2)	393	Largest(2)	31
Smallest(2)	41	Smallest(2)	14
Confidence Level(95,0%)	10,767	Confidence Level(95,0%)	0,523221

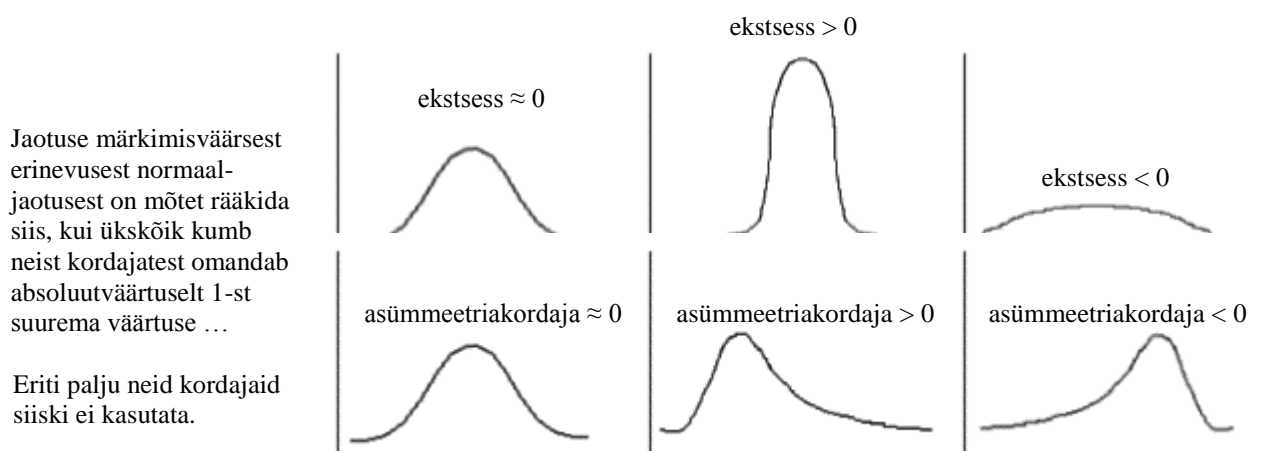
- Kas teate, mida kõik *Exceli* poolt arvutatud suurused enesest kujutavad?
- Kas ahvenate pikkuse ja massi jaotusi võib pidada sümmeetrilisteks? Miks?

- **Selgitus protseduuri *Descriptive Statistics* lisavalikutest eelmisel lehel:**

- valiku *Summary statistics* tulemusena arvutab *Excel* kaheteistkümne põhilise arvarakteristiku väärtused;
- valiku *Confidence Level for Mean: 95%* tulemusena arvutatakse suurus, mis tuleb keskmisele juurde liita või lahutada, saamaks ülemist ja alumist usalduspiiri; vaikumisi kasutatava 95% asemele võib ise trükkida mõne teise arvu (näiteks 90 või 99);
- valikute *Kth Largest* ja *Kth Smallest* tulemusena väljastatakse järjekorranumbriga K väärtus vastavalt suuremate ja väiksemate väärtuste poolt lugedes; $K = 1$ korral on tulemuseks maksimaalne ja minimaalne väärtus, et aga need suurused sisalduvad ka valiku *Summary statistics* väljundis, on antud juhul mõistlik tellida näiteks suuruselt järgmised väärtused (siis $K = 2$).

- **Uuritava tunnuse jaotuse kuju iseloomustamine**

Enamusest protseduuri *Descriptive Statistics* väljundis sisalduvatest arvarakteristikuteist on ennegi juttu olnud. Siiski on siin ka kaks uut suurust, mida kasutatakse peamiselt uuritava tunnuse jaotuse kuju iseloomustamiseks – need suurused on **ekstsess e järsakuskordaja** (ingl *kurtosis*) ja **asümmeetriakordaja** (ingl *skewness*). Sellest, mida need karakteristikud mõeldavad, annavad parema ettekujutuse järgnevad joonised:



- **Jaotuse sümmeetrilisuse üle otsustamisel kasutatakse sageli (asümmeetriakordaja asemel) keskmise ja mediaani võrdlust.**

Nimelt, kuna aritmeetiline keskmine on tundlik erandlike väärtuste suhtes, siis vihjab

$\bar{x} > med$ sellele, et jaotuse kuju on parempoolse ebasümmeetriaga (leiduvad üksikud teistest palju suuremad väärtused, ja seega asümmeetriakordaja > 0),

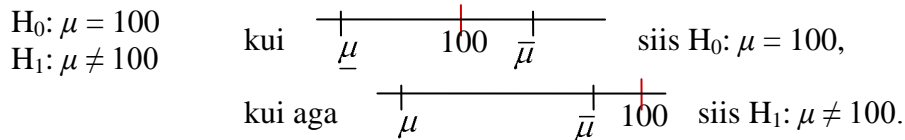
$\bar{x} < med$ aga sellele, et jaotuse kuju on vasakpoolse ebasümmeetriaga (leiduvad üksikud teistest palju väiksemad väärtused, ja seega asümmeetriakordaja < 0).

- 2) *** Selles alapunktis räägitavaist usalduspiiridest ja hüpoteeside testimisest tuleb juttu alles järgmises loengus, aga eelnevate statistikursuste olemasolul ei tohiks antud leheküljel presenteeritu üle mõistuse käia. ***

Kas olulisuse nivool 0,05 võib väita, et Emajõe vesikonna ahvenate populatsiooni keskmine mass erineb 100 grammist (eeldades, et uuritav andmestik kujutab enesest juhuslikku valimit vastavast populatsioonist)?

--- Lahenduse idee ja tööjuhend ---

Kui võrreldav konstant jääb keskmise 95%-usalduspiiride vahele, ei saa väita, et keskmine erineks konstandist, ja vastupidi:



	kaal		pikkus	
1070				
1071				
1072	Mean	118,6	Mean	20,44749
1073	Standard Error	5,4561	Standard Error	0,26514
1074	Median	95	Median	20
1075	Mode	75	Mode	18
1076	Standard Deviation	72,997	Standard Deviation	3,547326
1077	Sample Variance	5328,6	Sample Variance	12,58352
1078	Kurtosis	5,9845	Kurtosis	0,328945
1079	Skewness	2,034	Skewness	0,771968
1080	Range	483	Range	18
1081	Minimum	34	Minimum	14
1082	Maximum	517	Maximum	32
1083	Sum	21230	Sum	3660,1
1084	Count	179	Count	179
1085	Largest(2)	393	Largest(2)	31
1086	Smallest(2)	41	Smallest(2)	14
1087	Confidence Level(95,0%)	10,767	Confidence Level(95,0%)	0,523221
1088	Alumine usalduspiir	=B1072-B1087		
1089	Ülemine usalduspiir	=B1072+B1087		

Usalduspiirid keskmisele leitakse valemist $\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$

Excel väljastab toodud valemi mõlemad liidetavad, mille alusel on lihtne mõlemad usalduspiirid välja arvutada.

Alumine usalduspiir	107,84
Ülemine usalduspiir	129,37

Kumb väidetest on korrektne?

- Ei ole alust väita, et Emajõe vesikonna ahvenate populatsiooni keskmine mass erineks 100 grammist ($p > 0,05$).
- Emajõe vesikonna ahvenate populatsiooni keskmine mass erineb / on suurem kui 100 gr ($p < 0,05$).

- Kõige kiirem variant mõne olulisema arvarakteristiku väärtuse teada saamiseks *Excelis* (näiteks minimaalse ja maksimaalse väärtuse välja selgitamiseks andmete kontrollimise või sagedustabeli klasside moodustamise huvides) on järgmine:
 - võtate blokki huvipakkuvad lahtrid,
 - klikite hiire parempoolse klahviga *Exceli* tööakna alumisel ribal (nn olekuribal, vt joonist) ja valite avanenud rippmenüüst teid huvitavad funktsioonid (nt. *Min*),
 - valitud funktsiooni väärtus kuvatakse samas akna alumisel serval, aga seda ei trükiti kuhugi tabeli lahtrisse.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O			
1	kala nr	liik	ryhm	pyygikoht	sesoon	sugu	kaal	pikkus	arv	lõpuse arv	silmapar	arv	soolepar	arv	Diphy	lõpusepar	silmapar	soolepar
2	374	Ahven	roovkala	Võrtsjärv	kevad-suvi	i	131	22,5	10	22	26	0	1	1				
3	372	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	202	25										
4	560	Ahven	roovkala	Peipsi	sygis-talv	e	134	21,5										
5	377	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	84	18,5										
6	365	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	145	23										
7	335	Ahven	roovkala	Praaga	kevad-suvi	i	85	19,5	2									
8	64	Ahven	roovkala	Kastre	kevad-suvi	e	166	23	2									
9	527	Ahven	roovkala	Võrtsjärv	sygis-talv	i	152	22										
10	525	Ahven	roovkala	Võrtsjärv	sygis-talv	e	218	24,5										
11	376	Ahven	roovkala	Võrtsjärv	kevad-suvi	i	111	22										
12	375	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	132	23										
13	360	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	351	31	2									
14	347	Ahven	roovkala	Praaga	kevad-suvi	i	70	19	1									
15	344	Ahven	roovkala	Praaga	kevad-suvi	e	61	18	3									
16	343	Ahven	roovkala	Praaga	kevad-suvi	i	93	20	1									
17	310	Ahven	roovkala	Peipsi	kevad-suvi	e	87	19	1									
18	755	Ahven	roovkala	Peipsi	kevad-suvi	i	133	22	2									
19	754	Ahven	roovkala	Peipsi	kevad-suvi	i	194	23,5										
20	753	Ahven	roovkala	Peipsi	kevad-suvi	i	192	23,5										
21	521	Ahven	roovkala	Võrtsjärv	sygis-talv	i	229	24										
22	487	Ahven	roovkala	Kastre	sygis-talv	e	45	15										
23	373	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	141	21										
24	371	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	248	27										
25	361	Ahven	roovkala	Võrtsjärv	kevad-suvi	e	250	27,5	1									
26	342	Ahven	roovkala	Praaga	kevad-suvi	i	43	15,5	1									
27	341	Ahven	roovkala	Praaga	kevad-suvi	i	58	18	2									
28	339	Ahven	roovkala	Praaga	kevad-suvi	i	70	18,5	2									
29	334	Ahven	roovkala	Praaga	kevad-suvi	e	123	22	2									
30	329	Ahven	roovkala	Kastre	kevad-suvi	e	76	18	5									
31	319	Ahven	roovkala	Peipsi	kevad-suvi	e	80	18,5	7									
32	318	Ahven	roovkala	Peipsi	kevad-suvi	e	77	18,5										
33	316	Ahven	roovkala	Peipsi	kevad-suvi	e	67	18										
34	313	Ahven	roovkala	Peipsi	kevad-suvi	e	119	22,5										

OSA 2 --- Sagedustabelid ---

❖ Sagedustabel mitteamvulisele või diskreetsele arvtoonusele

Leidke hariliku laiussi noorvormiga nakatunud ja mitte nakatunud haugide arvud ja osakaalud.

--- Tööjuhend ---

The image illustrates the process of creating a PivotTable in Microsoft Excel. It shows the 'Create PivotTable' dialog box, the 'PivotTable Field List' task pane, and the 'Value Field Settings' dialog box. The final PivotTable shows the count of 'Diphylob' for each 'liik' category.

Create PivotTable Dialog Box:

- Choose the data that you want to analyze:
 - Select a table or range: data!\$A\$1:\$Q\$1043
 - Use an external data source: (not selected)
- Choose where you want the PivotTable report to be placed:
 - New Worksheet: (not selected)
 - Existing Worksheet: (selected)
 - Location: PivotTable!\$A\$20

PivotTable Field List Task Pane:

- Choose fields to add to report:
 - liik: (checked)
 - Diphylob: (checked)
- Drag fields between areas below:
 - Report Filter: (empty)
 - Column Labels: (empty)
 - Row Labels: liik
 - Values: Sum of Diphylob

Value Field Settings Dialog Box:

- Source Name: Diphylob
- Custom Name: Count of Diphylob
- Summarize Values By: Show Values As
- Summarize value field by: Count

PivotTable Data:

Row Labels	Count of Diphylob
Ahven	3
Haug	28
Koha	0
Latikas	0
Luts	59
Särg	0
Grand Total	90

Final PivotTable Data:

Row Labels	Count of Diphylob
Ahven	179
Haug	214
Koha	111
Latikas	173
Luts	137
Särg	228
Grand Total	1042

- Jätke tabelist välja ülejäänud kalaliigid peale haugi;
- jagage tabel veergudeks laiussi esinemise alusel (tunnus 'Diphylob');
- laiussi noorvormiga nakatunud ja mitte nakatunud haugide absoluutarvudele lisaks ka suhteliste sageduste leidmiseks lohistage tunnus 'Diphylob' ka teine kord väärtuste lahtrisse Σ Values ning määrake funktsiooniks Count ja esitusviisiks % of Row Total:

Row Labels	Column Labels	0	1	Grand Total
Haug				
Count of Diphylob		186	28	214
Count of Diphylob2		186	28	214
Total Count of Diphylob		186	28	214
Total Count of Diphylob2		186	28	214

Tulemus:

	A	B	C	D	E
16					
17					
18					
19					
20			Diphylob		
21	liik	Data	0	1	Grand Total
22	Haug	Count of Diphylob	186	28	214
23		Count of Diphylob2	86,92%	13,08%	100,00%
24		Total Count of Diphylob	186	28	214
25		Total Count of Diphylob2	86,92%	13,08%	100,00%
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					

PivotTable Field List

Choose fields to add to report:

- kala_nr
- liik
- ryhm
- pyygikoht
- sesoon
- sugu
- kaal
- pikkus
- arv_l6pusepar
- arv_silmapar
- arv_soolepar
- arv_Diphylob
- l6pusepar
- silmapar
- soolepar
- Diphylob
- par_per_kala

Drag fields between areas below:

Report Filter: [Empty]

Column Labels: Diphylob

Row Labels: liik, Σ Values

Values: Count of Diphylob, Count of Diphylob2

❖ **Sagedustabel arvtunnusele**

ÜLESANNE: konstrueerige sagedustabel haugide pikkusele.

--- **Tööjuhend** ---

- Pidevale arvtunnusele sagedustabeli tegemiseks on MS Excelis kolm moodust: funktsioon *FREQUENCY*, statistikaprotseduur *Histogram* ja *PivotTable*.

Nii funktsioon *FREQUENCY* kui ka statistikaprotseduur *Histogram* eeldavad, et kasutaja on eelnevalt välja mõelnud loodavad klassid ja sisestanud klasside ülemised piirid Exceli töölehele.

- Otsustamaks klasside arvu ja suuruse üle, peab esmalt omama ülevaadet vaatluste arvust ja uuritava tunnuse väärtuste ulatusest.

Seega – **leidke** mistahes viisil **haugide arv** ning nende **vähim** ja **suurim pikkus** (soovi korral võib haugide kohta käivaist andmeist teha koopia eraldi töölehele).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q					
1	kala_nr	liik	ryhm	pyygiikoht	sesoon	sugu	kaal	pikkus	arv	iguse	arv	silmapar	arv	soolepar	arv	Diphyllo	igusepar	silmapar	soolepar	Diphyllob	par_per	kala
2	291	Haug	roovkala	Peipsi	kevad-suvi	e	1071	52	22	39	2	4	1	1	1	1	1	1	1	1	1	6
3	380	Haug	roovkala	Võrtsjärv	kevad-suvi	i	602	45	50	4	5	0	1	1	1	1	1	1	1	1	0	6
4	230	Haug	roovkala	Praaga	sygis-talv	i	1108	55	52	18	3	0	1	1	1	1	1	1	1	1	0	5
5	179	Haug	roovkala	Praaga	sygis-talv	i	775	50	16	4	36	2	1	1	1	1	1	1	1	1	1	5
6	555	Haug	roovkala	Peipsi	sygis-talv	i	555	45,5	30	4	2	0	1	1	1	1	1	1	1	1	0	5
7	477	Haug	roovkala	Kärevere	sygis-talv	e	883	51,5	20	1	13	0	1	1	1	1	1	1	1	1	0	5
8	351	Haug	roovkala	Võrtsjärv	kevad-suvi	i	541	46	32	0	1	0	1	0	1	0	1	0	1	0	0	3
9	281	Haug	roovkala	Peipsi	kevad-suvi	e	1163	55	28	3	6	0	1	1	1	1	1	1	1	1	0	5
10	280	Haug	roovkala	Peipsi	kevad-suvi	i	717	49	46	4	3	0	1	1	1	1	1	1	1	1	0	6
11	198	Haug	roovkala	Kärevere	sygis-talv	i	819	50	26	0	5	0	1	0	1	0	1	0	1	0	0	4

- Vaatluste arv 214 annab vihje, et haugid võiks pikkuse järgi jagada $\sqrt{214} \approx 15$ klassi (see ei ole absoluutne tõde, kui klasside piiridega ja/või sisuliste järeldustega paremini sobib, võib väärtused jagada ka väiksemasse arvu klassidesse).
- Teadmise, et 214 haugi pikkused varieeruvad 26-st sentimeetrist 74 sentimeetrini, alusel võiks moodustada 13 4-sentimeetrist klassi: 25-28; 29-32; 33-36; ...; 69-72; 73-76.

NB! Toodud tähistus sobib üksnes täisarvuliste väärtuste korral (miks?), matemaatiliselt korrektne on esitada klassid poollõikudena: (24,28], (28,32], (32,36], ..., (68,72] ja (72,76].

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

- Märkusi:

– mistahes abitabeli ja algse andmetabeli vahele on soovitatav jätta vähemalt üks tühi veerg (või rida) – **miks?** (kui ei tea, küsi!);

– Excel tõlgendab ette antud väärtusi järgmiselt:

'28' ↔ ' ≤ 28 ';

'32' ↔ ' > 28 ja ≤ 32 '

jne (ehk igasse klassi kuuluvaks loetakse need väärtused, mis on

väiksemad või võrdsed ette antud piirist ja mis ei kuulu eelnevatesse klassidesse);

– viimast klassi ' 76 '=(72,76] ette ei anta, sest sagedustabeli moodustamisel teeb Excel ise täiendava klassi, kuhu loeb kokku kõik eelnevatesse klassidesse mitte kuuluvad väärtused.

T
pikkus_klassid
28
32
36
40
44
48
52
56
60
64
68
72

▪ Funktsioon **FREQUENCY**

Kõige kiirem variant lasta Excelil kokku lugeda, kui palju vaatlusi mingisse ette antud klassi kuulub, on kasutada funktsiooni **FREQUENCY** (kui statistikaprotseduuride pakett *Data Analysis* ei ole teie arvutis installeeritud, aitab just see funktsioon hädast välja).

Erinevalt enamusest MS Exceli funktsioonidest on funktsioon **FREQUENCY** **massiivifunktsioon**, st et selle funktsiooni tulemuseks ei puugi olla üks väärtus eelnevalt valitud lahtris, vaid hulk väärtusi eelnevalt valitud lahtriteblokkis.

T	U
pikkus_klassi	Sagedused
28	
32	
36	
40	
44	
48	
52	
56	
60	
64	
68	
72	

- Esimese asjana peale klassipiiride Exceli töölehele sisestamist tuleb võtta blokki lahtrid töölehel kohas, kuhu soovitakse sagedusi arvutada; arvutatavate sageduste ja seeläbi blokki võetavate lahtrite arv on määratud konstrueeritava sagedustabeli klasside arvuga (üks täiendav blokki võetud lahter vastab Exceli poolt täiendavalt moodustatavale klassile).

- Trükite selekteeritud lahtriblokki (koheselt sellesse lahtrisse, millest blokki võtmist alustasite, uuesti klikkida esimesel lahtril ei tohi!!)

=FREQUENCY(H2:H215;T2:T13)

misjärel vajutate alla klahvid **'Shift'** ja **'Ctrl'** ning seejärel **'Enter'** (st. 3 klahvi korraga).

pikkus_klassi	Sagedused
28	=FREQUENCY(H2:H215;T2:T13)
32	FREQUENCY(data_array; bins_array)
36	
40	
44	
48	
52	
56	
60	
64	
68	
72	

- Alternatiiv taolisele funktsiooni klaviatuurilt sisestamisele on lisada funktsioon menüüsid ja abiaknaid kasutades:

Tulemus:

S	T	U
pikkus_klassi	Sagedused	
28		1
32		0
36		2
40		5
44		37
48		62
52		68
56		25
60		7
64		3
68		1
72		2
		1

▪ **Protseduur Histogram**

Data-sakk → Data Analysis... → Histogram

Ütleb Excelile, et ette antud lahtri-blokkide esimeses reas on nimed

Väljundtabeli (vasaku ülemise nurga) asukoht

Lisaks tavalistele sagedustele arvutatakse ka kumulatiivsed suhtelised sagedused e. jaotus

pikkus_klassid	Frequency	Cumulative %
28	1	0,47%
32	0	0,47%
36	2	1,40%
40	5	3,74%
44	37	21,03%
48	62	50,00%
52	68	81,78%
56	25	93,46%
60	7	96,73%
64	3	98,13%
68	1	98,60%
72	2	99,53%
More	1	100,00%

NB! Lahtri 'Bin Range' võib jätta ka tühjaks – siis moodustab Excel klassid ise.

PROOVIGE!

▪ **PivotTable**

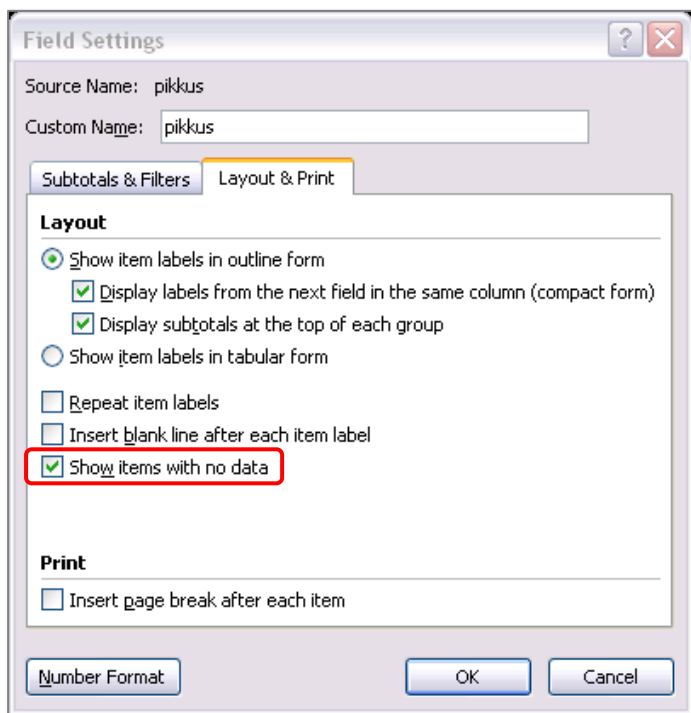
NB! *PivotTable*'i abil saab pidevale tunnusele teha sagedustabeli üksnes siis, kui tunnuse kõik väärtused on mõõdetud (andmebaasis pole „auke“).

Row Labels	Count of kala_nr	Count of kala_nr2
26	1	0,47%
34	2	0,93%
38	1	0,47%
39	1	0,47%
40	3	1,40%
41	2	0,93%
41,5	1	0,47%
42	7	3,27%
42,5	2	0,93%
43	11	5,14%
43,5	3	1,40%
44	11	5,14%
44,5	1	0,47%
45	13	6,07%
45,5	4	1,87%
46	11	5,14%
46,5	1	0,47%
47	12	5,61%
47,5	5	2,34%
48	15	7,01%
48,5	1	0,47%
49	18	8,41%
49,5	1	0,47%
50	13	6,07%
50,5	1	0,47%
51	18	8,41%
51,5	2	0,93%
52	14	6,54%
52,5	1	0,47%
53	7	3,27%
54	4	1,87%
55	7	3,27%
56	6	2,80%
57	1	0,47%

Row Labels	Count of kala_nr	Count of kala_nr2
24-28	1	0,47%
32-36	2	0,93%
36-40	2	0,93%
40-44	29	13,55%
44-48	58	27,10%
48-52	69	32,24%
52-56	33	15,42%
56-60	13	6,07%
60-64	2	0,93%
64-68	1	0,47%
68-72	3	1,40%
72-76	1	0,47%
Grand Total	214	100,00%

Jama on see, et *Excel* vaikimisi ei näita tühje klasse (näiteks kõrval olevast tabelist on puudu klass 28-32).

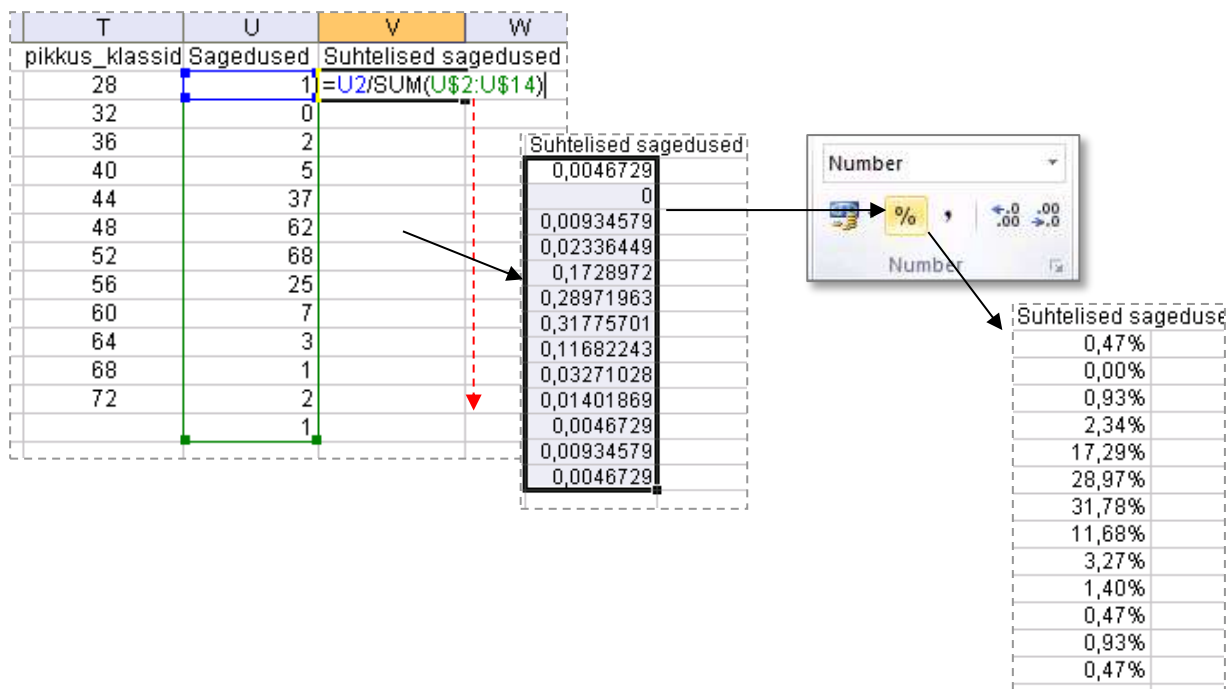
Lahendus:



Tulemus:

Row Labels	Count of kala_nr	Count of kala_nr2
<24		0,00%
24-28	1	0,47%
28-32		0,00%
32-36	2	0,93%
36-40	2	0,93%
40-44	29	13,55%
44-48	58	27,10%
48-52	69	32,24%
52-56	33	15,42%
56-60	13	6,07%
60-64	2	0,93%
64-68	1	0,47%
68-72	3	1,40%
72-76	1	0,47%
>76		0,00%
Grand Total	214	100,00%

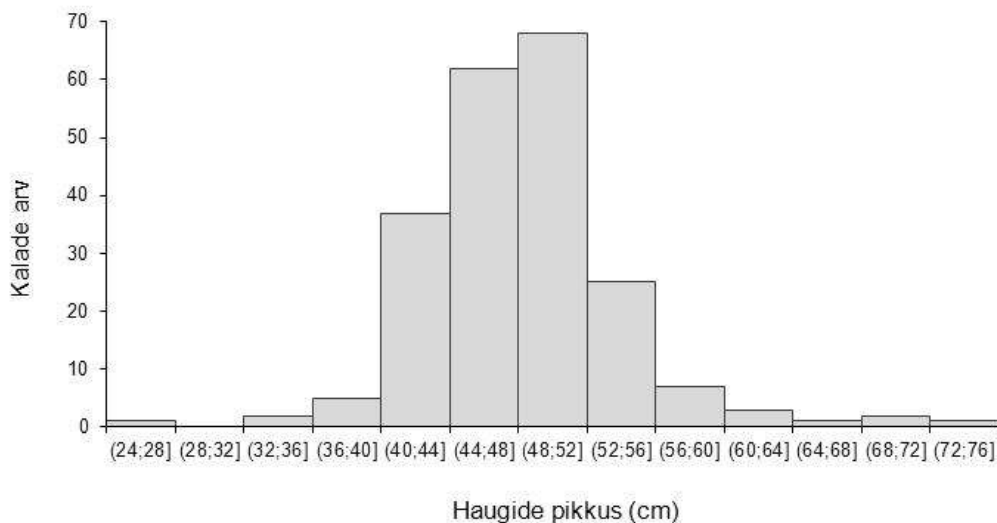
- Arvutage suhtelised sagedused (%-des) ka funktsiooniga *FREQUENCY* või protseduuriga *Histogram* konstrueeritud sagedustabeli juurde ja illustreerige sagedustabelit histogrammiga.



Pikkused	pikkus_klassid	Sagedused	Suhtelised sagedused
(24;28]	28	1	0,47%
(28;32]	32	0	0,00%
(32;36]	36	2	0,93%
(36;40]	40	5	2,34%
(40;44]	44	37	17,29%
(44;48]	48	62	28,97%
(48;52]	52	68	31,78%
(52;56]	56	25	11,68%
(56;60]	60	7	3,27%
(60;64]	64	3	1,40%
(64;68]	68	1	0,47%
(68;72]	72	2	0,93%
(72;76]		1	0,47%

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

Näpuharjutuseks püüdke joonis kujundada selliseks, nagu kõrvaloleval pildil.



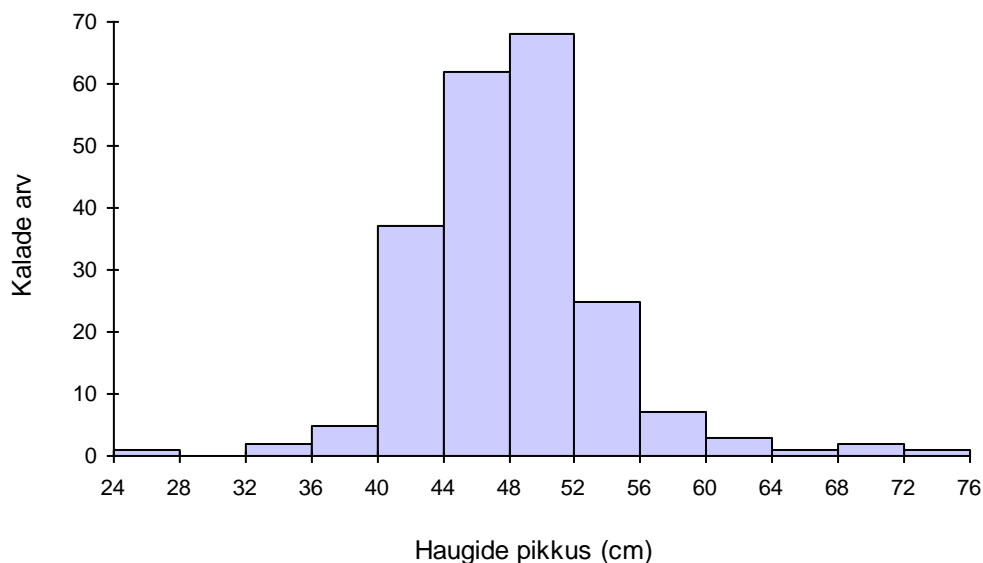
OSA 3

--- Pisut joonistest ---

Näiteid vähe keerukamatest *Exceli* joonistest võib leida aadressilt http://www.eau.ee/~ktanel/joonised_excelis/.

❖ Histogramm

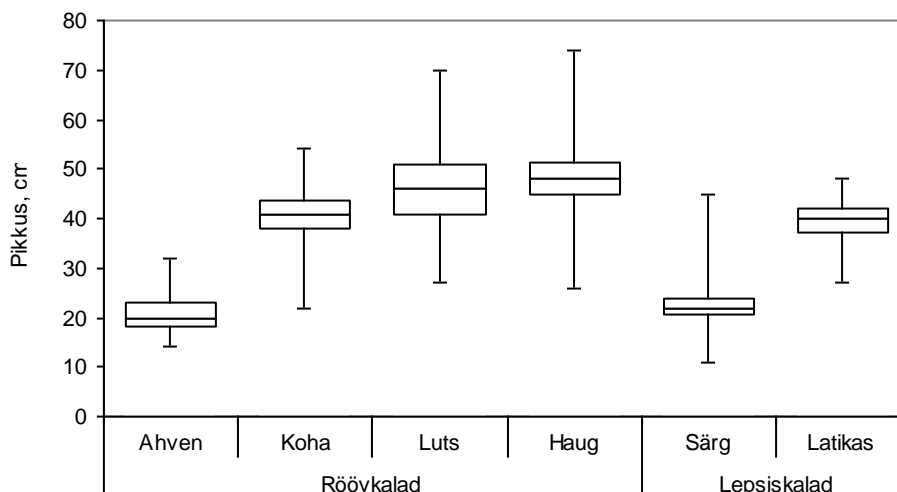
Eelmisel lehel esitatud histogramm matemaatiliselt korrektselt kujul (pideva arvteljena esitatud x-teljega) võiks olla järgmine:



Juhendi taolise pideva arvteljena esitatud x-teljega histogrammi tarvis leiate Internetist aadressilt http://ph.emu.ee/~ktanel/joonised_excelis/joonis1.php; lehelt <http://www.treeplan.com/better.htm> on alla laaditav ka tasuta lisamoodul *Excelile* taolise joonise tegemiseks.

❖ Karp-vurrud diagramm

Ülesanne: illustreerige erinevat liiki kalade pikkuseid karp-vurrud diagrammiga (inglisekeeles *box plot*) kujul



Probleem: Exceli joonisetüüpide hulgas antud diagrammi ei ole.

Lahendus:

- arvutada kalaliikide kaupa mediaan, kvartiilid ning minimaalne ja maksimaalne väärtus,
- konstrueerida nende alusel abitabel joonisel esitatavate andmetega,
- joonistada tulpdiagramm ning
- modifitseerida viimast, saamaks joonisele soovitud väljanägemist.

Tööjuhend.

1) Esimese sammuna tuleb konstrueerida abitabel ning arvutada sinna karp-vurrud diagrammi joonistamiseks vajalikud arvkarakteristikud (miinimum, alumine kvartiil, mediaan, ülemine kvartiil, maksimum), ja seda iga kalaliigi tarvis.

Selleks tuleks andmestik sorteerida kalaliigi järgi (**NB!** alternatiivset varianti on kirjeldatud lehekülje lõpus) ning rakendada arvkarakteristikute arvutamiseks vastavaid funktsioone (kõik viis joonise tegemiseks vajalikku suurust on leitavad näiteks funktsiooniga QUARTILE.INC (vanemais Exceli versioonides funktsioon QUARTILE), mille teise argumenti väärtus määrab ära arvutatava suuruse: 0 vastab miinimumile, 1 alumisele kvartiilile jne).

Seejuures on kiireim variant

- arvutada välja näiteks mediaan, fikseerides ära argumentidena ette antavad lahtrid (st pannes lahtri aadressis dollarimärgid rea- ja veerutähiste ette – ahvena puhul näiteks kujul \$H\$2:\$H\$180),
- kopeerida valemit sisaldav lahter ka teistesse veergudesse ning muuta ära vaid funktsiooni QUARTILE.INC viimane argument (vt järgmine joonis),
- järgmise kalaliigi tarvis tuleb protseduuri korrata.

Tulemus:

			Mediaan	25%-punkt	75%-punkt	Min	Max
Röövkalad	Ahven		20	18	23	14	32
	Koha		41	38	43,5	22	54
	Luts		46	41	51	27	70
	Haug		48,25	45	51,5	26	74
Lepsiskalad	Särg		22	20,5	24	11	45
	Latikas		40	37	42	27	48

NB! Veel kiirem viis on rakendada funktsiooni QUARTILE koos funktsiooniga IF ja seejuures võib ka funktsiooni QUARTILE teise argumenti anda ette väärtusena vastavas lahtris:

Eelneval joonisel kontrollib funktsiooni QUARTILE (või QUARTILE.INC) esimeseks argumentiks olev funktsioon IF esmalt

(1), kas andmetabeli järjekordses reas oleva kala liik (lahtrites B2:B1043) võrdub loodava abitabeli kala liigi veerus oleva väärtusega (mis sisaldub lahtris T3),

(2), kui võrdub, võtab funktsioon QUARTILE arvutuste aluseks vastavas reas paikneva kala pikkuse, kui aga ei võrdu, saab funktsioon QUARTILE argumentiks tühja väärtuse (""), mille ta lihtsalt kõrvale jätab;

(3) funktsiooni QUARTILE teine argument, mis näitab, milline kvartiil arvutada, võetakse abitabeli peale täiendavalt lisatud reast.

Et taolise valemi näol on tegu **massiivifunktsiooniga**, tuleb selle rakendamiseks vajutada alla klahvid 'Ctrl'+ 'Shift' ja alles siis 'Enter'.

Toodud kujul korra sisestatud valem on juba kopeeritav kõigisse teistesse abitabeli lahtritesse ega vaja enam ühtki muutmist. Ka andmetabeli eelnev kalaliikide järgi sorteerimine ei ole vajalik!

2) Järgmise sammuna tuleb leitud arvarakteristikute tabeli alusel konstrueerida uus, Excelis loodavale tulpdiagrammile aluseks olev tabel:

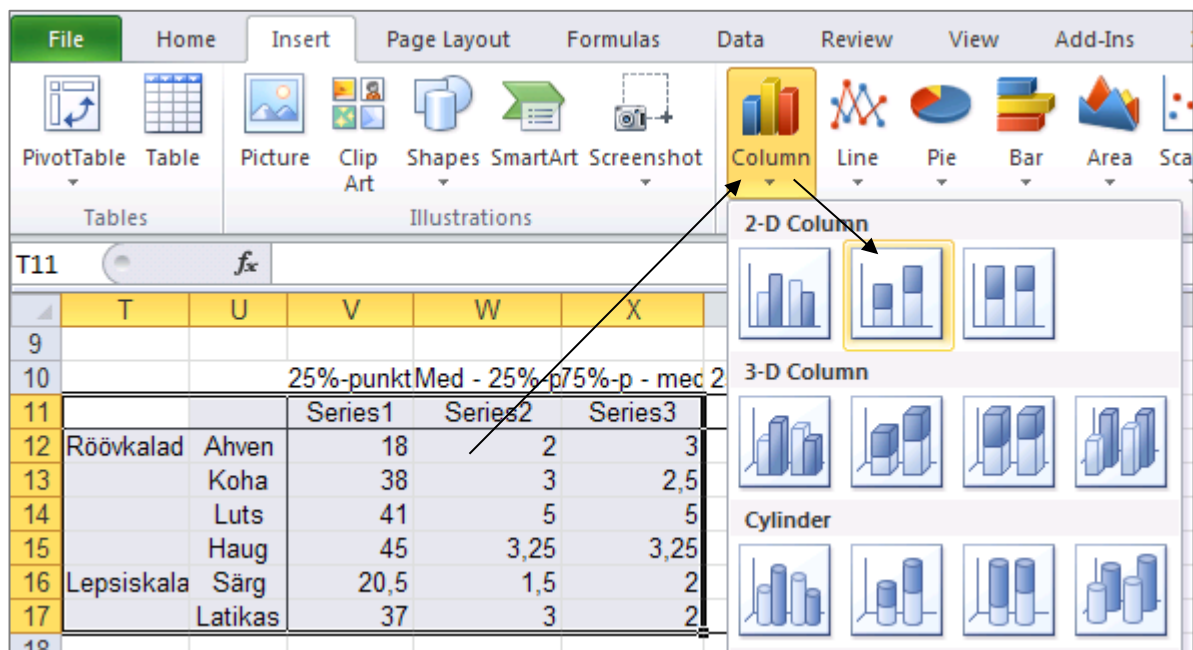
		25%-punkt	Med – 25%-p	75%-p – med	25%-p – min	Max – 75%-p
		Series1	Series2	Series3	Error1	Error1
Röövkalad	Ahven	18	2	3	4	9
	Koha	38	3	2,5	16	10,5
	Luts	41	5	5	14	19
	Haug	45	3,25	3,25	19	22,5
Lepsiskalad	Särg	20,5	1,5	2	9,5	21
	Latikas	37	3	2	10	6

Veerud uues tabelis on püütud nimetada vastavalt nendes olevate arvude rollile loodavas joonises.

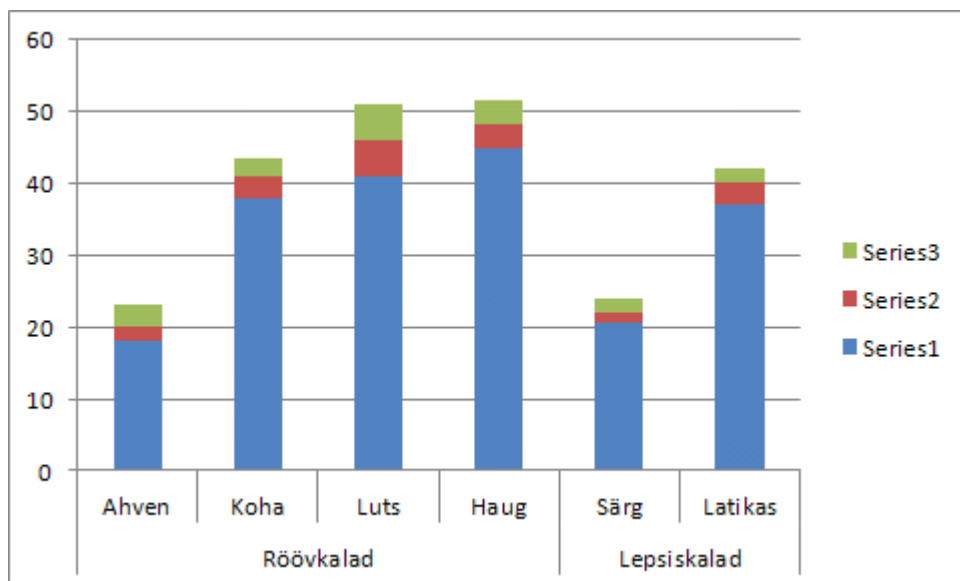
Seejuures

- 'Series1' = 'Alumine kvartiil' (need arvud vastavad esimese tulba kõrgusele tulpdiagrammil e karbi alumisele servale karp-vurrud diagrammil);
- 'Series2' väärtused on arvutatud kui 'Mediaan – Alumine kvartiil' (see suurus määrab ära teise tulba kõrguse tulpdiagrammil e mediaanile vastava joone karp-vurrud diagrammil);
- 'Series3' väärtused on arvutatud kui 'Ülemine kvartiil – Mediaan' (see suurus määrab ära kolmanda tulba kõrguse tulpdiagrammil e karbi ülemise serva karp-vurrud diagrammil);
- 'Error bars1' väärtused on arvutatud kui 'Alumine kvartiil – Miinimum' (selle suuruse alusel joonistatakse 'Series1'-le vastavale tulpale alumine veajoon, mis karp-vuurud diagrammi mõistes tähendab joont alumisest kvartiilist minimaalse väärtuseni);
- 'Error bars2' väärtused on arvutatud kui 'Maksimum – Ülemine kvartiil' (selle suuruse alusel joonistatakse 'Series3'-le vastavale tulpale ülemine veajoon, mis karp-vuurud diagrammi mõistes tähendab joont ülemisest kvartiilist maksimaalse väärtuseni).

3) Uue abitabeli veergude 'Series1'-'Series3' alusel tuleb konstrueerida tulpdiagramm, kus tabeli veerud vastaksid erinevatele andmeseriiatele



Tulemus:



4) Saadud joonisel tuleb

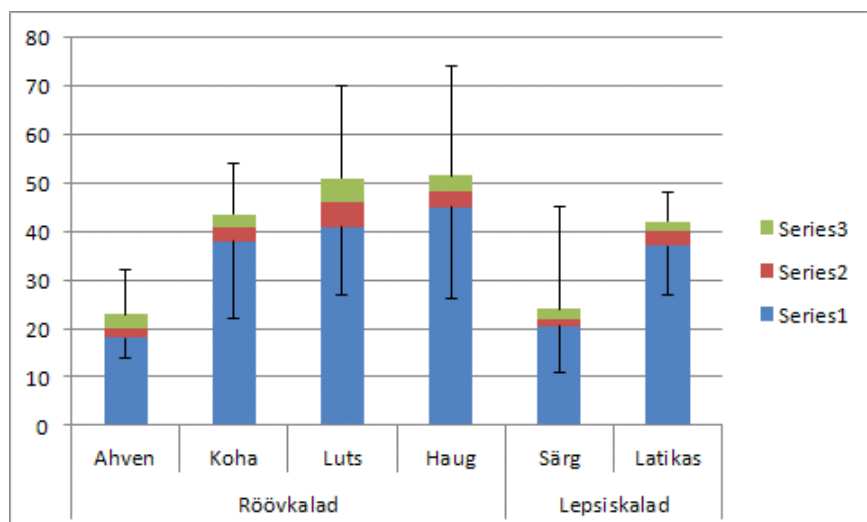
- lisada 'Series1' tulbale alumine veajoon veerus 'Error1' olevate väärtuste alusel,

The screenshot shows an Excel spreadsheet with data for 'Röövkalad' and 'Lepsiskalad'. A stacked bar chart is displayed below the data. The chart has three series: Series1 (blue), Series2 (red), and Series3 (green). The 'Error1' column in the data table contains values for Series1. The 'Format Error Bars' dialog box is open, showing 'Vertical Error Bars' with 'Minus' direction and 'Cap' end style. The 'Custom Error Bars' dialog box is also open, showing the formula for the negative error value: '=data!\$Y\$1'.

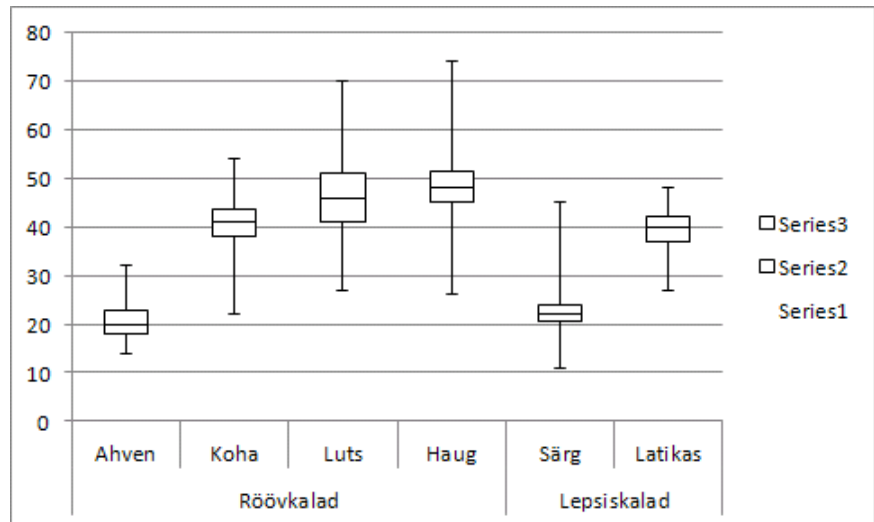
		Mediaan	25%-punkt	75%-punkt	Min	Max
Röövkalad	Ahven	20	18	23	14	32
	Koha	41	38	43,5	22	54
	Luts	46	41	51	27	70
	Haug	48,25	45	51,5	26	74
Lepsiskalad	Särg	22	20,5	24	11	45
	Latikas	40	37	42	27	48

		25%-punkt	Med - 25%-p	75%-p - med	25%-p - min	Max - 75%-p
Röövkalad	Ahven	18	2	3	4	9
	Koha	38	3	2,5	16	10,5
	Luts	41	5	8	14	19
	Haug	45	3,25	3,25	19	22,5
Lepsiskalad	Särg	20,5	1,5	2	9,5	21
	Latikas	37	3	2	10	6

- analoogselt lisada tulbale 'Series3' ülemine veajoon veerus 'Error2' olevate väärtuste järgi; tulemus peaks olema järgmine:



- kaotada ära nii sisu kui ka piirjooned 'Series1' tulbalt ning sisu 'Series2' ja 'Series3' tulpadelt (lisades vajadusel viimastele piirjooned);
tulemus:



- viimase etapina tuleks
 - kaotada ära legend ja horisontaalsed ruudujooned,
 - lisada telgedele nimed,
 - vähendada tulpade (karp-vurrud diagrammide) vahelist kaugust (*Gap Width*), st et vähendada tühja ebainformatiivset ala joonisel,
 - soovi korral ümbritseda telgedega piiratud diagrammiala joonega ning kaotada ära kogu joonist ümbritsev joon (nii on joonis selgem!) ning
 - muuta joonise proportsioone.

Tulemus:

