

Praks 2

Eel- ja järeltöö

1. Salvestage arvutisse andmestik lammas.xls (http://ph.emu.ee/~ktanel/VL_1112/lammas.xls).
2. Avage salvestatud fail *MS Excel*'is.
- ...
- ...
3. Peale ülesannete lahendamist salvestage fail nimega 'perekonnanimi_lammas.xls' ja saatke e-meiliga aadressil tanel.kaart@emu.ee.

Kommentaariid andmestiku kohta

Rümpade EUROP klassifitseerimine on EL riikides kehtiv lihakehade klassifitseerimissüsteem, kus hinnatakse iga lihakeha kommertsväärtust.

Rümbad jaotatakse järgmistesse kategooriatesse:

- alla 12 kuu vanuste lammaste e tallede rümbad (tähis „L“) ning kõigi ülejäänud lammaste rümbad (tähis „S“);
- visuaalselt hinnatud lihakusklasside osas eristatakse: „E“ (ekstra), „U“ (väga hea), „R“ (hea), „O“ (rahuldav), „P“ (lahja), „P-“ (eriti lahja);
- visuaalselt hinnatud rasvasusklasside osas eristatakse: „1“ (väherasvane), „2“ (kergelt rasvane), „3“ (keskmiselt rasvane), „4“ (rasvane), „5“ (väga rasvane).

Antud andmestik sisaldab 686 lambarümba andmeid (56-lt omanikult) 2002. aasta sügisest (tapetud ja hinnatud kõik samas tapamajas).

Iga lamba kohta on lisaks eelnevalt nimetatud kolmele rühmitavale tunnusele:

'rümba kategooria' $\in \{L, S\}$, 'lihakusklass' $\in \{E, U, R, O, P, P-\}$ ja 'rasvasusklass' $\in \{1, 2, 3, 4, 5\}$, fikseeritud ka see,

kas loom oli pärit jõudluskontrollialusest karjast või mitte (vastavalt 'jkk' = 1 või 0),

kas realiseerimine leidis aset läbi ELaS-i turustusgrupi või mitte (vastavalt 'realis' = 1 või 0),

samuti on teada rümba mass (kg) ja hind (EEK), mille alusel on arvutatud rümba 1 kg hind (EEK/kg).

Ülesanded

Praktikumi tehniline pool hõlmab peamiselt *Pivot Table*'i ja diagrammide kasutamist *MS Excel*'is, lisaks ka veel χ^2 -testi ja regressioonianalüüsi.

1. Kirjeldage lammaste jagunemist EUROP klassifitseerimissüsteemi alusel, leides erinevatesse klassidesse kuuluvate rümpade arvud ja protsendid (seda siis 3 tunnuse tarvis – rümba üldkategooria, lihakusklass ja rasvasusklass).
 - Kui mõnda lihakus- ja/või rasvasusklassi kategooriat esineb väga vähe, pange see kokku sarnase naaberkategooriaga.
 - Illustreerige saadud tabelleid sektordiagrammidega, kirjutades igale sektorile juurde sellele vastava väärtuse ja esinemise suhtelise sageduse protsentides.
2. Kas rümpade jagunemine rasvasusklassidesse sõltub rümba üldkategooriast?
 - Võimaliku seose kirjeldamiseks konstrueerige (uuele töölehele) vastav 2-mõõtmeline sagedustabel, viimasesse leidke nii rea- kui ka veeruprotsendid ja sõnastage lause(d), kasutades vähemalt kahte leitud suhtelistest sagedustest.
3. Jätkuna punktile 2 testige rümpade üldkategooriatesse ja rasvasuklassidesse jagunemise vahelise seose statistilist olulisust.
 - Et oleks selge, mida te üldse testite, pange esmalt kirja kontrollitav hüpoteeside paar.
 - Järgnevalt konstrueerige uus 2-mõõtmeline sagedustabel, mis sisaldab üksnes absoluutseid sagedusi, selle alusel arvutage tunnuste sõltumatuse juhule (nullhüpoteesile) vastavad sagedused ja
 - teostage funktsiooni `CHISQ.TEST` (*Excel* 2003-s `CHITEST`) abil χ^2 -test – viimane võrdleb empiirilisi (andmetabelist arvatud) sagedusi teoreetiliste (sõltumatuse juhule vastavate) sagedustega ja väljastab olulisuse tõenäosuse p väärtuse.
 - Sõnastage lõppjärelendus (viidates sõnastuses ka p -väärtusele, millel järelendus baseerub).
4. Prognoosige tallerümpade 1 kg hinda lähtuvalt rümba massist. Kui palju võinuks 2002. aasta sügisel keskmiselt raha saada 20 kg kaaluva tallerümba eest.
 - ❖ Esmalt sorteerige/filtreerige algandmed vastavalt rümpade üldkategooriale ja tehke uuele töölehele koopia tallerümpade massidest ja 1 kg hindadest.
 - ❖ Teostage regressioonianalüüs graafiliselt.
 - ❖ Selleks laske *Excel*il joonistada hajuvusdiagramm (punktdiagramm), kus x -teljel paiknevad rümpade massid ja y -teljel hinnad.
 - ❖ Valmis diagrammile lisage regressioonisirge, regressioonivõrrand ja viimase baasil saadavate prognooside täpsust kirjeldav determinatsioonikordaja R^2 .

- ❖ Lisaks tavalisele lineaarsele regressioonanalüüsile sobitage punktiparvest läbi ka ruutfunktsiooni graafik ning tellige sellegi tarvis *Excel*'lt võrrand ja R^2 (parema võrdlemise huvides värvige vastav joon ja parameetrid näiteks punaseks).
- ❖ Kumba seost – lineaarset või ruutseost – eelistada tallerümba 1 kg hinna prognoosimisel? Miks?
- Pange töölehele kirja regressioonivõrrand ja prognoosige 20 kg kaaluva tallerümba hinda.

5. Lisaülesanne. Leidke *Pivot Table*'i abil uuele töölehele rümpade arv, keskmine, minimaalne ja maksimaalne mass ning massi standardhälve sõltuvalt lamba päritolust (jõudluskontrolli alusest karjast või mitte).

- Illustreerige leitud keskmisi tulpdiagrammiga, kus rümba masside varieeruvust kirjeldavad standardhälbed on kujutatud nõ veajoontena (joonise tegemiseks tehke vajalikest *Pivot Table*'i abil leitud väärtustest abitabel).
- Sorteerige (või filtreerige) algandmed vastavalt jõudluskontrolli alla kuulumisele ning tehke leitud keskmistega samale lehele abitabel, mis sisaldab ühes veerus jõudluskontrollialusest karjast pärit rümpade masse ja teises veerus mitte jõudluskontrollialusest karjast pärit rümpade masse.
- Testige keskmiste masside erinevuse statistilist olulisust (esmalt F-test ja selle tulemusest lähtuvalt õiget tüüpi t-test). Sõnastage lõppjärelus.

Kui aru ei saa (näiteks, mida mingi funktsioon teeb või miks midagi just näidatud kujul tööjuhendis realiseeritud on), siis küsi!

Illustreeritud (ja mittetäielik) tööjuhend

1. Konstrueerime järgnevalt näitena sagedustabeli rümba lihakusklassi kohta, analoogselt käib sagedustabelite tegemine ka rümba üldkategoriale ja rasvasusklassile.

❖ Paigutage kursor andmetabeli suvalisse lahtrisse → *Insert*-sakk → *PivotTable* →

Choose the data that you want to analyze

Select a table or range
Table/Range: andmed!\$A\$1:\$I\$687

Use an external data source
Choose Connection...

Connection name:

Choose where you want the PivotTable report to be placed

New Worksheet
 Existing Worksheet

Location: andmed!\$L\$2

Loodava tabeli vasaku ülemise nurga asukoht

	A	B	C	D	E	F	G	H	I	J	K	L
1	Lammas	Jkk	Realis	R_mass	R_kat	Lihakusklass	Rasvasusklass	R_hind	R_kg_hind			
2	142656	1	1	23,8	L	R	3	928,2	39			
3	142687	1	1	15,2	L	R	3	577,6	39			
4	142366	1	1	15,6	L	R	2	624	40			
5	142335	1	1	24,4	L							
6	142465	1	1	21,1	L							
7	142564	1	1	19,1	L							
8	142472	1	1	22,5	L							
9	142267	1	1	21,7	L							
10	142298	1	1	23,8	L							
11	142571	1	1	23,2	L							
12	142670	1	1	18,1	L							
13	142311	1	1	14,4	L							
14	142588	1	1	12,6	L							
15	142373	1	1	20,5	L							
16	160735	1	1	22,5	L							
17	52870	1	1	18,9	L							
18	52849	1	1	13,4	L							
19	142380	1	1	21,1	L							
20	142397	1	1	19,9	L							
21	52658	1	1	14,2	L							

Drop Column Fields Here

Drop Row Fields Here

Drop Value Fields Here

Tulemus:

Drag fields between areas below:

Report Filter

Column Labels

Row Labels

Values

Lihakusklass

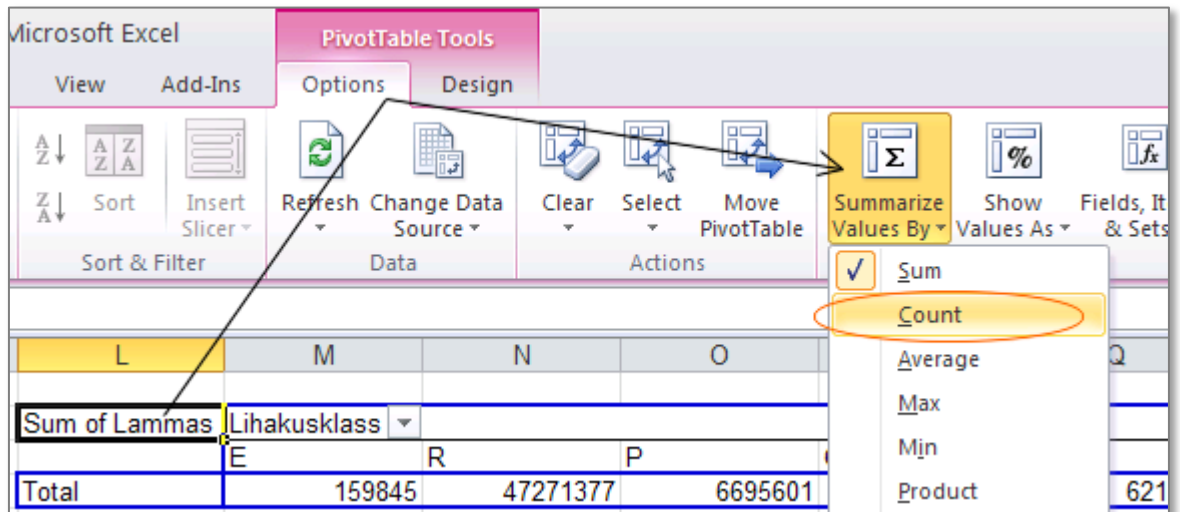
Sum of Lammas

Loodava tabeli vasaku ülemise nurga asukoht

Tulemuseks saadud sagedustabel:

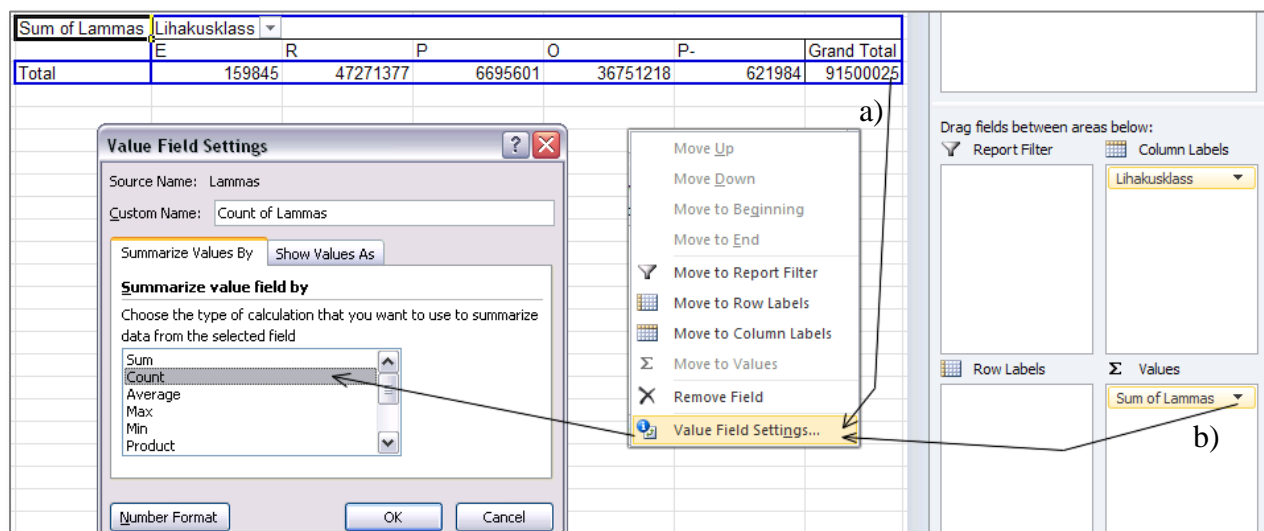
Sum of Lammas	Lihakusklass							
	E	R	P	O	P-			Grand Total
Total	159845	47271377	6695601	36751218	621984			91500025

- ❖ Vaikimisi arvutab *Excel* lammaste numbrite summa ...
Et selle asemel lihtsalt kokku lugeda, kui mitu lammast mingisse lihakusklassi kuulus, tuleb ära muuta *Pivot Table*-s kasutatav funktsioon (*Sum* asemel *Count*):
PivotTable Tools-sakk → *Options* → *Summarize Values By* →



Alternatiivina võib *Pivot Table*-s rakendatavat funktsiooni muuta ka

- klikkides tabelil hiire parempoolse nupuga või
- klikkides *Pivot Table* konstrueerimise aknas lahtris *Values* muuta soovitava funktsiooni järel paikneval kolmnurgal:



- ❖ Lihakusklasside sisuliselt õiges järjekorras esitamiseks (*Excel* sorteerib tähestikulises, mitte sisulises järjekorras) on lihtsaim variant vales kohas olev klass lihtsalt ümber tõsta (klikkides selleks klassi nimel (näiteks lahtril 'P') ja tõstes lahtri servast kinni hoides õigesse kohta):

Count of Lammas	Lihakusklass						Grand Total
	E	R	P	O	P-		
Total	1	347	53	277	8	686	

Tulemus:

Count of Lammas	Lihakusklass						Grand Total
	E	R	O	P	P-		
Total	1	347	277	53	8	686	

- ❖ Lisaks absoluutsetele sagedustele võiks leida ka suhtelised sagedused.

a) Lohistage *PivotTable Field List*'s tunnus 'Lammas' ka teine kord lahtrisse *Values*;

b) nõudke, et *Excel* jagaks tabeli erinevate funktsioonide alusel ridadeks, mitte veergudeks (lohistage kastike Σ *Values* lahtrisse *Row Labels*);

c) määrake vajadusel ka uue rea tarvis funktsiooniks *Count* (*Sum* asemel) ning

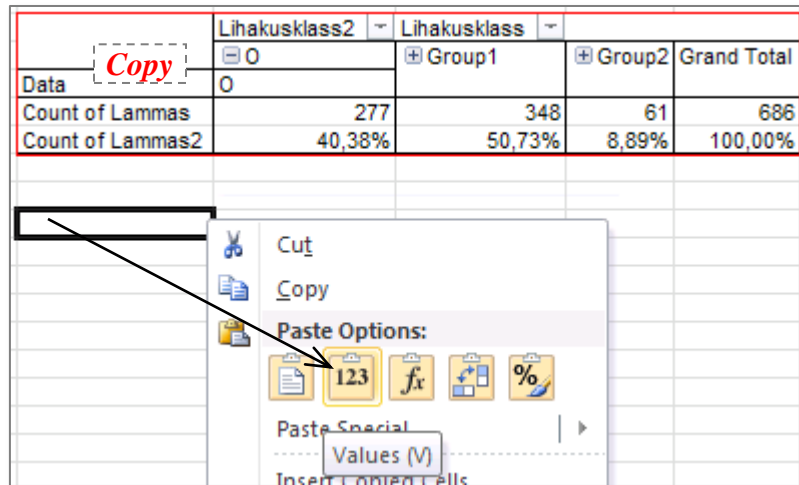
d) nõudke väärtuste esitamist protsentidena (*PivotTable Tools*-sakk \rightarrow *Options* \rightarrow *Show Values As* \rightarrow *% of Row Total*).

Tulemus:

	Lihakusklass						Grand Total
Data	E	R	O	P	P-		
Count of Lammas	1	347	277	53	8	686	
Count of Lammas2	0,15%	50,58%	40,38%	7,73%	1,17%	100,00%	

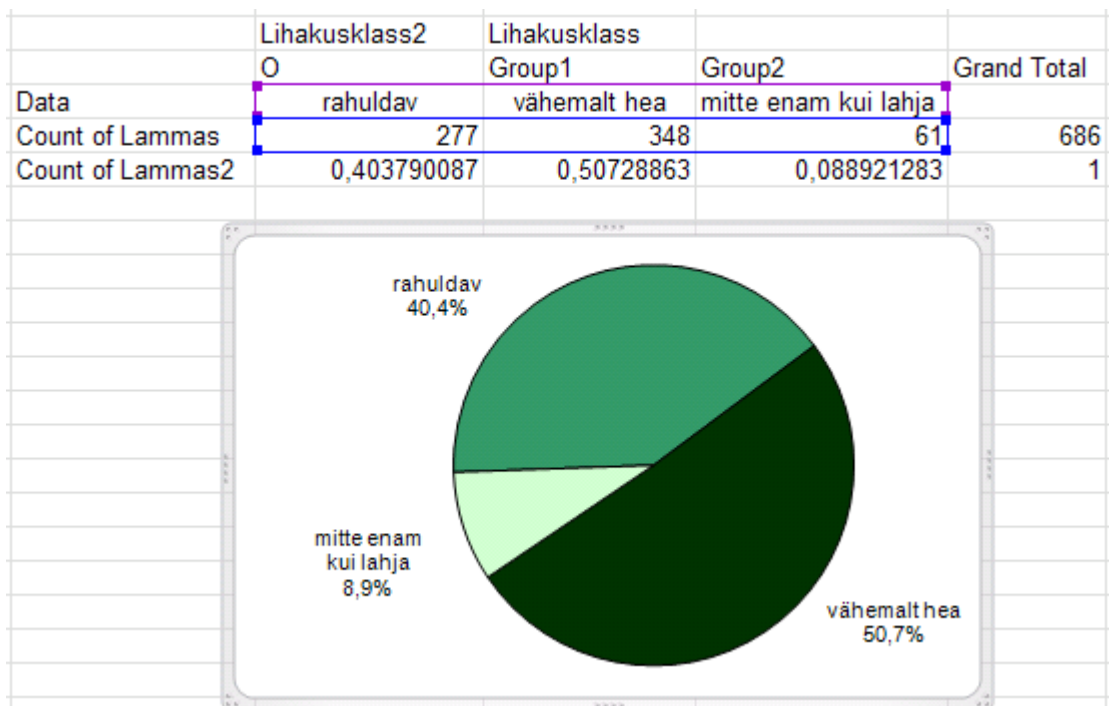
- ❖ *Pivot Table*'i abil konstrueeritud tabeli põhjal kenade jooniste tegemiseks on sageli mõttekas teha vajalikest väärtustest abitabel ja joonistada diagramm abitabeli alusel.

Põhjuseks on see, et otse *Pivot Table* alusel joonise tegemise tulemuseks on nn *Pivot Chart*, mis on sarnaselt *Pivot Table*'ga lingitud andmetabeliga, seeläbi kergesti täiendatav ja ümberarvutatav, aga ei võimalda muuta kõike tavalisel *Excel*'i diagrammil muudetavat (või on see märksa keerulisem).



Järgnevalt andke lihakusklassidele sisuliselt õiged nimed – trükkige need moodustatud abitabelisse, ja konstrueerige abitabeli vastavate lahtrite alusel sektordiagramm.

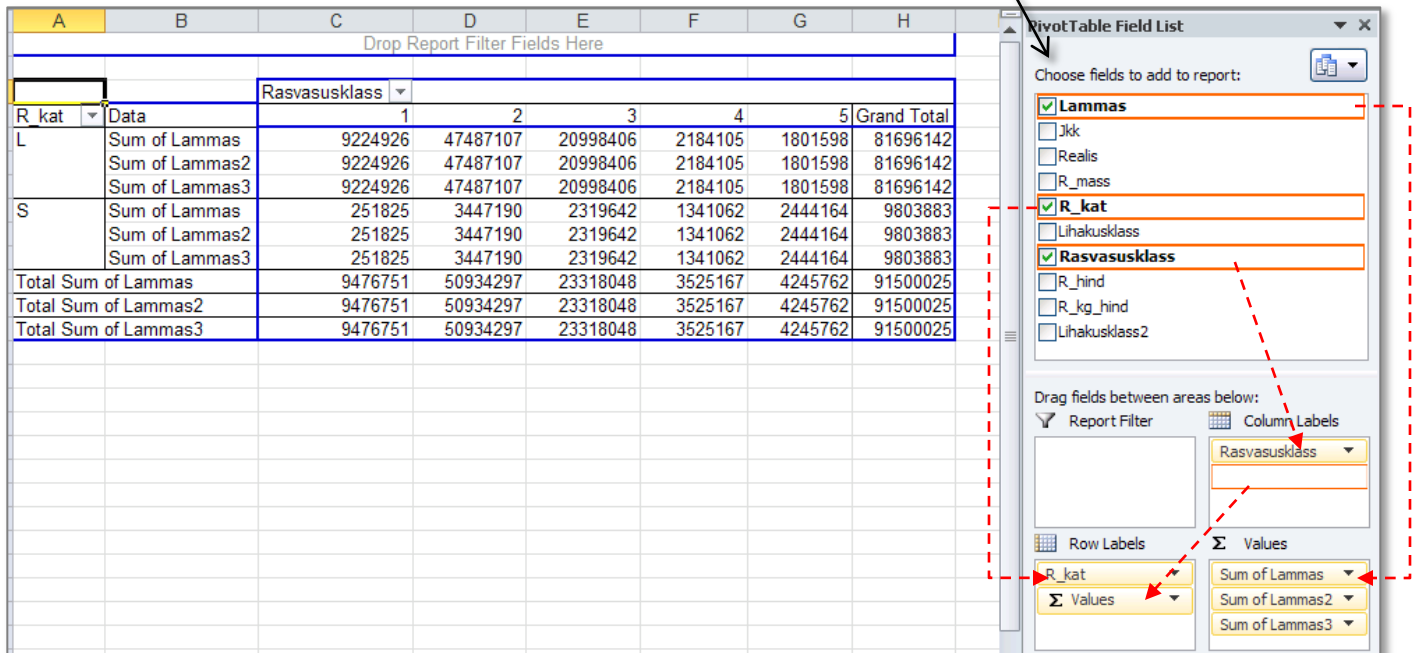
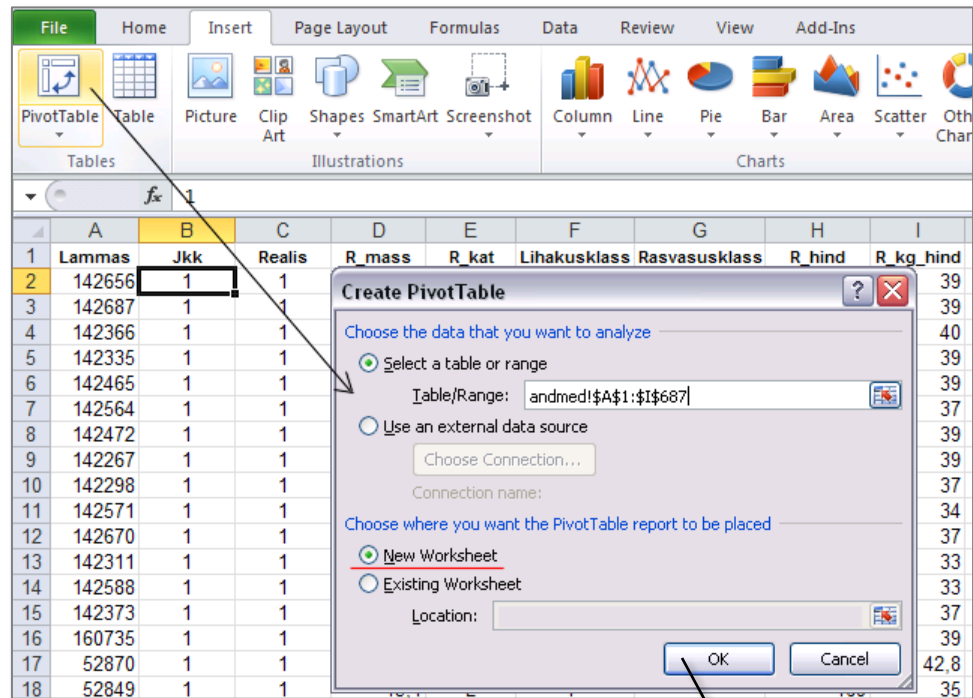
Kujundage saadud joonis (lisage sektoritele vastavate lihakusklasside nimed ja esinemissagedused, muutke soovi korral värve):



- ❖ Järgnevalt konstrueerige sagedustabelid ja joonistage nende alusel sektordiagrammid ka rümba (üld)kategoriale ja rasvasusklassile.

2. Kas rümpade jagunemine rasvasusklassidesse sõltub rümba üldkategorias?

Võimaliku seose kirjeldamiseks konstrueerige (uuele töölehele) vastav 2-mõõtmeline sagedustabel, viimasesse leidke nii rea- kui ka veeruprotsendid ja sõnastage lause(d), kasutades vähemalt kahte leitud suhtelistest sagedustest.



Muutke kõik kolm *Exceli* poolt leitud summat vaatluste arvudeks (*Sum* → *Count*):

R kat	Data	1	2	3	4	5	Grand Total
L	Sum of Lammas	69	301	132	14	10	526
	Sum of Lammas2	69	301	132	14	10	526
	Sum of Lammas3	69	301	132	14	10	526
S	Sum of Lammas	2	35	58	34	31	160
	Sum of Lammas2	2	35	58	34	31	160
	Sum of Lammas3	2	35	58	34	31	160
Total Sum of Lammas		71	336	190	48	41	686
Total Sum of Lammas2		71	336	190	48	41	686
Total Sum of Lammas3		71	336	190	48	41	686

R kat	Data	1	2	3	4	5	Grand Total
L	Count of Lammas	69	301	132	14	10	526
	Count of Lammas2	69	301	132	14	10	526
	Count of Lammas3	69	301	132	14	10	526
S	Count of Lammas	2	35	58	34	31	160
	Count of Lammas2	2	35	58	34	31	160
	Count of Lammas3	2	35	58	34	31	160
Total Count of Lammas		71	336	190	48	41	686
Total Count of Lammas2		71	336	190	48	41	686
Total Count of Lammas3		71	336	190	48	41	686

Ning esitage 2. ja 3. vaatluste arv (*Count of ... 2 ja 3*) vastavalt rea- ja veeruprotsendina:

R kat	Data	1	2	3	4	5	Grand Total
L	Count of Lammas	69	301	132	14	10	526
	Count of Lammas2	69	301	132	14	10	526
	Count of Lammas3	69	301	132	14	10	526
S	Count of Lammas	2	35	58	34	31	160
	Count of Lammas2	2	35	58	34	31	160
	Count of Lammas3	2	35	58	34	31	160
Total Count of Lammas		71	336	190	48	41	686
Total Count of Lammas2		71	336	190	48	41	686
Total Count of Lammas3		71	336	190	48	41	686

Tulemus:

R kat	Data	1	2	3	4	5	Grand Total
L	Count of Lammas	69	301	132	14	10	526
	Count of Lammas2	13,12%	57,22%	25,10%	2,66%	1,90%	100,00%
	Count of Lammas3	97,18%	89,58%	69,47%	29,17%	24,39%	76,68%
S	Count of Lammas	2	35	58	34	31	160
	Count of Lammas2	1,25%	21,88%	36,25%	21,25%	19,38%	100,00%
	Count of Lammas3	2,82%	10,42%	30,53%	70,83%	75,61%	23,32%
Total Count of Lammas		71	336	190	48	41	686
Total Count of Lammas2		10,35%	48,98%	27,70%	7,00%	5,98%	100,00%
Total Count of Lammas3		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Kommentaariid.

(sõnastage lause(d), kasutades vähemalt kahte leitud suhtelistest sagedustest)

3. Rümpade üldkategoriatesse ja rasvasuklassidesse jagunemise vahelise seose statistiline olulisus.

- ❖ Et oleks selge, mida te üldse testite, pange esmalt kirja kontrollitav hüpoteeside paar.
- ❖ Järgnevalt konstrueerige uus 2-mõõtmeline sagedustabel, mis sisaldab üksnes absoluutseid sagedusi,

H₀ (nullhüpotees): ...
 H₁ (alternatiivne hüpotees): ...

Empiirilised (andmetest arvatud) sagedused

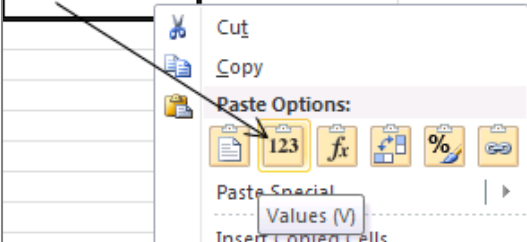
Count of Lammas	Rasvasuklass					Grand Total
R_kat	1	2	3	4	5	
L	69	301	132	14	10	526
S	2	35	58	34	31	160
Grand Total	71	336	190	48	41	686

- ❖ tehke konstrueeritud tabeli väärtustest koopia

Empiirilised (andmetest arvatud) sagedused

Count of Lammas	Rasvasuklass					Grand Total
R_kat	1	2	3	4	5	
L	69	301	132	14	10	526
S	2	35	58	34	31	160
Grand Total	71	336	190	48	41	686

Copy



- ❖ ja kustutage kopeeritud tabeli sisu (alles jätke rea- ja veerusummad!)

Count of Lammas	Rasvasuklass					Grand Total
R_kat	1	2	3	4	5	
L						526
S						160
Grand Total	71	336	190	48	41	686

Delete

- ❖ Arvutage uude tabelisse kustutatud sageduste asemele tunnuste sõltumatuse juhule (nullhüpoteesile) vastavad nõ teoreetilised sagedused (kujul: reasumma × veerusumma / vaatluste arv).

Teoreetiliste sageduste arvutamine:

	A	B	C	D	E	F	G
25							
26	Teoreetilised (nullhüpoteesile e sõltumatuse juhule vastavad) sagedused						
27	Count of Lammas	Rasvasusklass					
28	R_kat	1	2	3	4	5	Grand Total
29	L	=G29*B\$31/G\$31					526
30	S						160
31	Grand Total	71	336	190	48	41	686

... ..

	A	B	C	D	E	F	G
25							
26	Teoreetilised (nullhüpoteesile e sõltumatuse juhule vastavad) sagedused						
27	Count of Lammas	Rasvasusklass					
28	R_kat	1	2	3	4	5	Grand Total
29	L	54,44023324	257,6326531	145,6851312	36,80466472	31,43731778	526
30	S	16,55976676	78,36734694	44,3148688	11,19533528	=G30*F\$31/G\$31	160
31	Grand Total	71	336	190	48	41	686

- ❖ Teostage funktsiooni CHISQ.TEST (Excel 2003-s CHITEST) abil χ^2 -test – viimane võrdleb empiirilisi (andmetabelist arvatud) sagedusi teoreetiliste (sõltumatuse juhule vastavate) sagedustega ja väljastab olulisuse p väärtuse

Et oleks lihtsam aru saada, milliste arvude võrdlemisel χ^2 -test baseerub (ehk siis millised tabelite osad tuleb Excel'i vastavale funktsioonile ette anda), võib vastavad lahtrid selguse mõttes näiteks ära värvida.

Järgnevalt, nagu funktsioonide puhul ikka, tuleb kursor panna lahtrisse, kuhu soovitakse tulemust saada (ja juurde võiks enne ka kirjutada, mida arvutama hakatakse).

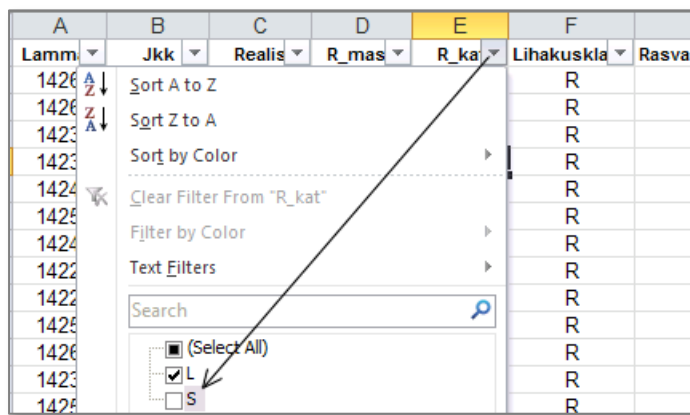
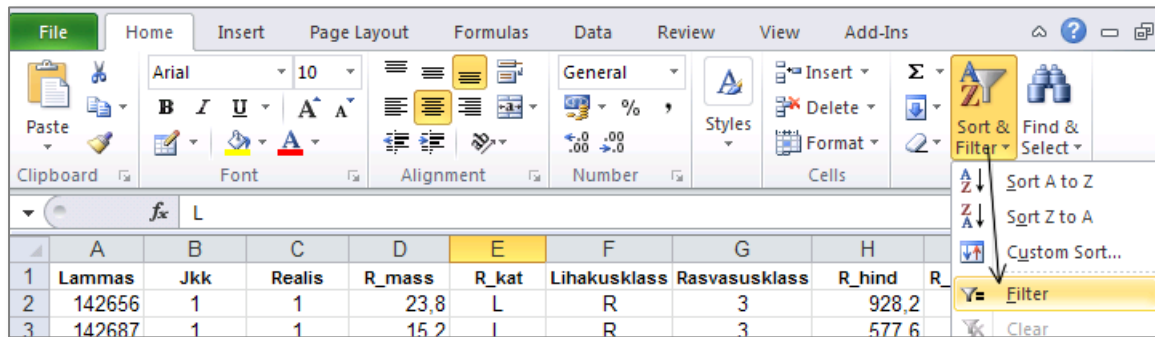
	A	B	C	D	E	F	G	
17								
18	Empiirilised (andmetest arvatud) sagedused							
19	Count of Lammas	Rasvasusklass						
20	R_kat	1	2	3	4	5	Grand Total	
21	L	69	301	132	14	10	526	
22	S	2	35	58	34	31	160	
23	Grand Total	71	336	190	48	41	686	
24								
25								
26	Teoreetilised (nullhüpoteesile e sõltumatuse juhule vastavad) sagedused							
27	Count of Lammas	Rasvasusklass						
28	R_kat	1	2	3	4	5	Grand Total	
29	L	54,44023324	257,6326531	145,6851312	36,80466472	31,43731778	526	
30	S	16,55976676	78,36734694	44,3148688	11,19533528	9,562682216	160	
31	Grand Total	71	336	190	48	41	686	
32								
33								
34	Hii-ruut-test	=CHISQ.TEST(B21:F22;B29:F30)						

- ❖ Sõnastage lõppjärgeldus (Kas seos on statistiliselt oluline? Miks te nii otsustasite?).

Hii-ruut-test 3,69291E-37

4. Prognoosige tallerümpade 1 kg hinda lähtuvalt rümba massist. Kui palju võinuks 2002. aasta sügisel keskmiselt raha saada 20 kg kaaluva tallerümba eest.

- ❖ Esmalt sorteerige/filtreerige algandmed vastavalt rümpade üldkategoriale ja tehke uuele töölehele koopia tallerümpade massidest ja 1 kg hindadest.

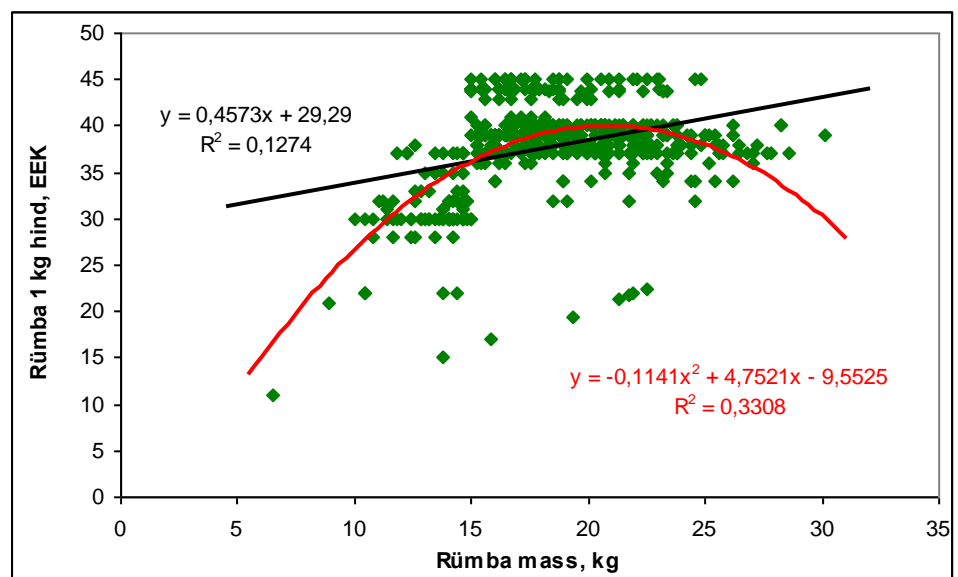
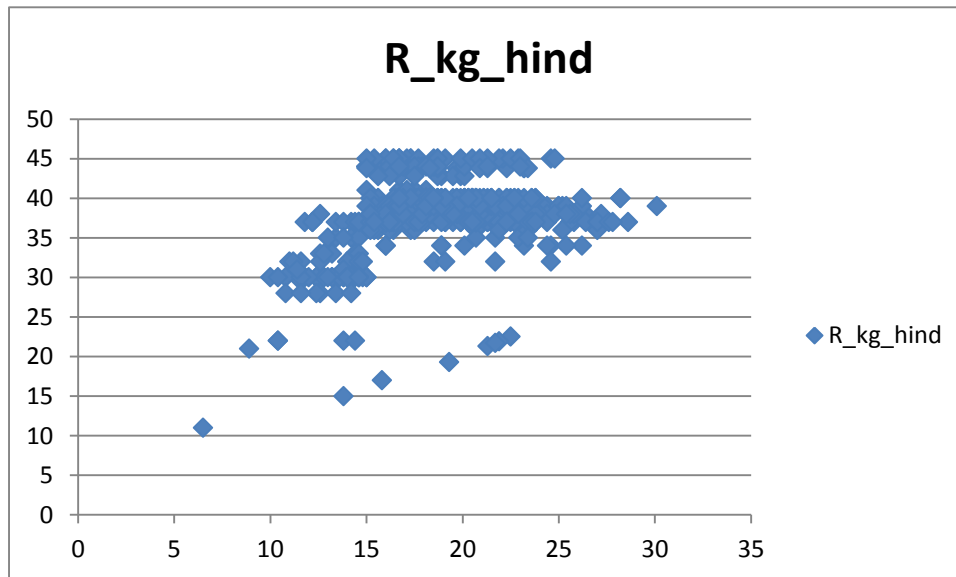


- ❖ ja tehke uuele töölehele koopia tallerümpade massidest ja 1 kg hindadest.

	A	B	C	D	E	F	G	H	I
1	Lamm	Jkk	Realis	R_mass	R_kat	Lihakusklass	Rasvasusklass	R_hind	R_kg hind
2	142656	1	1	23,8	L	R	3	928,2	39
3	142687	1	1	15,2	L	R	3	577,6	39
4	142366	1	1	15,6	L	R	2	624	40
5	142335	1	1	24,4	L	R	3	951,6	39
6	142465	1	1	21,1	L	R	3	822,9	39
7	142564	1	1	19,1	L	R	4	706,7	37
8	142472	1	1	22,5	L	R	3	877,5	39
9	142267	1	1	21,7	L	R	3	846,3	39
10	142298	1	1	23,8	L	R	4	880,6	37
11	142571	1	1	23,2	L	R	5	788,8	34
12	142670	1	1	18,1	L	R	4	669,7	37
13	142311	1	1	14,4	L	R	2	475,2	33
14	142588	1	1	12,6	L	R	2	415,8	33

❖ Teostage regressioonanalüüs graafiliselt.

- Selleks laske *Excelil* joonistada hajuvusdiagramm (punktdiagramm), kus x -teljel paiknevad rümpade massid ja y -teljel hinnad.
- Valmis diagrammile lisage regressioonisirge, regressioonivõrrand ja viimase baasil saadavate prognooside täpsust kirjeldav determinatsioonikordaja R^2 .
- Lisaks tavalisele lineaarsele regressioonanalüüsile sobitage punktisarvest läbi ka ruut-funktsiooni graafik ning tellige sellegi tarvis *Excel'*lt võrrand ja R^2 (parema võrdlemise huvides värvige vastav joon ja parameetrid näiteks punaseks).
- Kumba seost – lineaarset või ruutseost – eelistada tallerümba 1 kg hinna prognoosimisel? Miks?



❖ Pange töölehele kirja regressioonivõrrand ja prognoosige 20 kg kaaluva tallerümba hinda.