

Praktikum 2

Salvestage kursuse kodulehelt oma arvutisse andmestik *lammas.xls* (http://ph.emu.ee/~ktanel/VL_1112/lammas.xls).

Kommentaari andmestiku kohta

Lammaste andmebaas on moodustatud aastal 2003 ja selles andmebaasis on samas tapamajas tapetud ja hinnatud 686 lamba (56-lt omanikult) rümpade andmed. Igale rümbale on määratud

- rümba kategooria väärtustega L ja S (vastavalt alla 12 kuu vanuste lammaste e tallede rümbad ja kõigi ülejäänud lammaste rümbad),
- lihakusklass väärtustega E, U, R, O, P ja P- (tegu on EUROP klassifitseerimisega, mis on EL riikides kehtiv lihakehade klassifitseerimissüsteem, kus hinnatakse iga lihakeha kommerts-väärtust: E – ekstra, U – väga hea, R – hea, O – rahuldav, P – lahja, P- – eriti lahja),
- rasvasusklass väärtustega 1 kuni 5 (1 – väherasvane, 2 – kergelt rasvane, 3 – keskmiselt rasvane, 4 – rasvane ja 5 – väga rasvane).

Lisaks on fikseeritud ka see,

- kas loom oli pärit jõudluskontrollialusest karjast või mitte (vastavalt 1 või 0),
- kas realiseerimine leidis aset läbi ELaS-i turustusgrupi või mitte (vastavalt 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 *PivotTable* ja diagrammide kasutamist MS Excelis, lisaks ka veel χ^2 -testi, regressioonianalüüsi ja t-testi.

1. Kirjeldage lammaste jagunemist EUROP klassifitseerimissüsteemi alusel, leides erinevatesse klassidesse kuuluvate rümpade arvud ja protsendid (seda siis kolme tunnuse tarvis – rümba üldkategorია, lihakusklass ja rasvasusklass).
 - Kui mõnda lihaku- ja/või rasvasusklassi kategooriat esineb väga vähe, pange see kokku sarnase naaberkategooriaga.
 - Illustreerige saadud tabeleid sektordiagrammidega, kirjutades igale sektorile juurde sellele vastava väärtuse ja esinemise suhtelise sageduse protsentides.
2. Kas rümpade jagunemine rasvasusklassidesse sõltub rümba üldkategoriasest?
 - Võimaliku seose kirjeldamiseks konstrueerige (uuele töölehele) vastav kahemõõtmeline sagedustabel, viimasesse leidke nii rea- kui ka veeruprotsendid ja sõnastage lause(d), kasutades vähemalt kahte leitud suhtelistest sagedustest.
 - Testige rümpade üldkategoriasest ja rasvasusklassidesse jagunemise vahelise seose statistilist olulisust.
 - Konstrueerige uus kahemõõtmeline sagedustabel, mis sisaldab üksnes absoluutseid sagedusi, selle alusel arvutage tunnuste sõltumatuse juhule (nullhüpoteesile) vastavad sagedused ja
 - teostage funktsiooni `CHI.TEST` abil χ^2 -test – viimane võrdleb empiirilisi (andmetabelist arvutatud) sagedusi teoreetiliste (sõltumatuse juhule vastavate) sagedustega ja väljastab olulisuse tõenäosuse p väärtuse.
 - Sõnastage lõppjäreldus (viidates sõnastuses ka p -väärtusele, millel järeldus baseerub).
3. 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.
 - Teostage regressioonianalüüs 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 regressioonianalüüsile sobitage punktiparvest läbi ka ruut-funktsiooni graafik ning tellige sellegi tarvis Excelilt 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.

ÜLESANNE 1.

--- Sagedustabel *PivotTable*'ga ---

Kirjeldage lammaste jagunemist EUROP klassifitseerimissüsteemi alusel, leides erinevatesse klassidesse kuuluvate rümpade arvud ja protsendid (seda siis kolme tunnuse tarvis – rümba üldkategooria, lihakusklass ja rasvasusklass).

1. Konstrueerime järgnevalt näitena sagedustabeli rümba lihakusklassi kohta, analoogselt käib sagedustabelite tegemine ka rümba üldkategooriale 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 where you want the PivotTable report to be placed

☐ New Worksheet

☒ Existing Worksheet

Location: andmed!\$L\$2

Loodava tabeli vasaku ülemise nurga asukoht

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

Sum of Lammas

PivotTable Field List

Choose fields to add to report:

Lammas

Jkk

Realis

R_mass

R_kat

Lihakusklass

Rasvasusklass

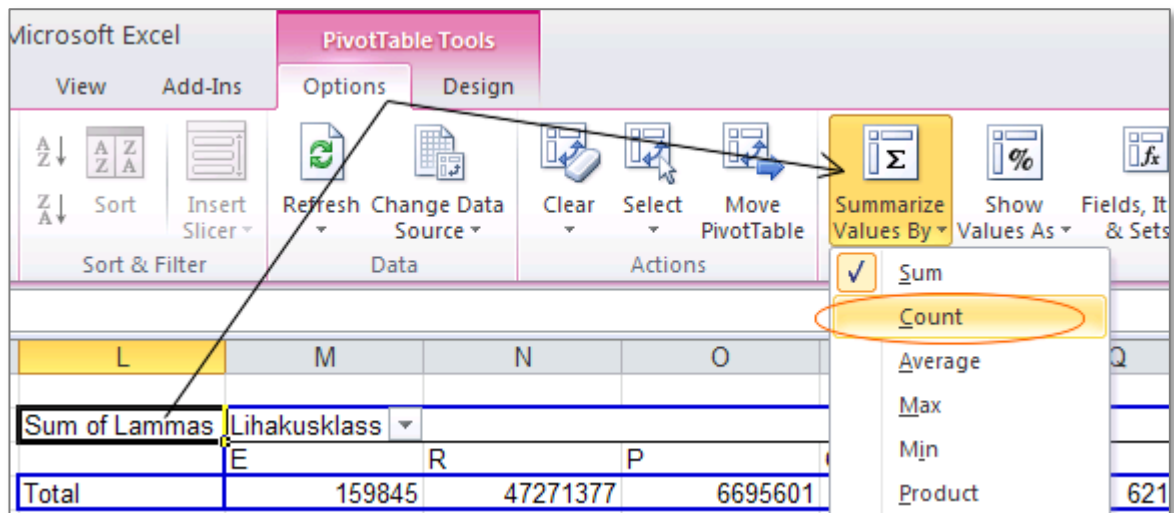
R_hind

Tulemuseks saadud sagedustabel:

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

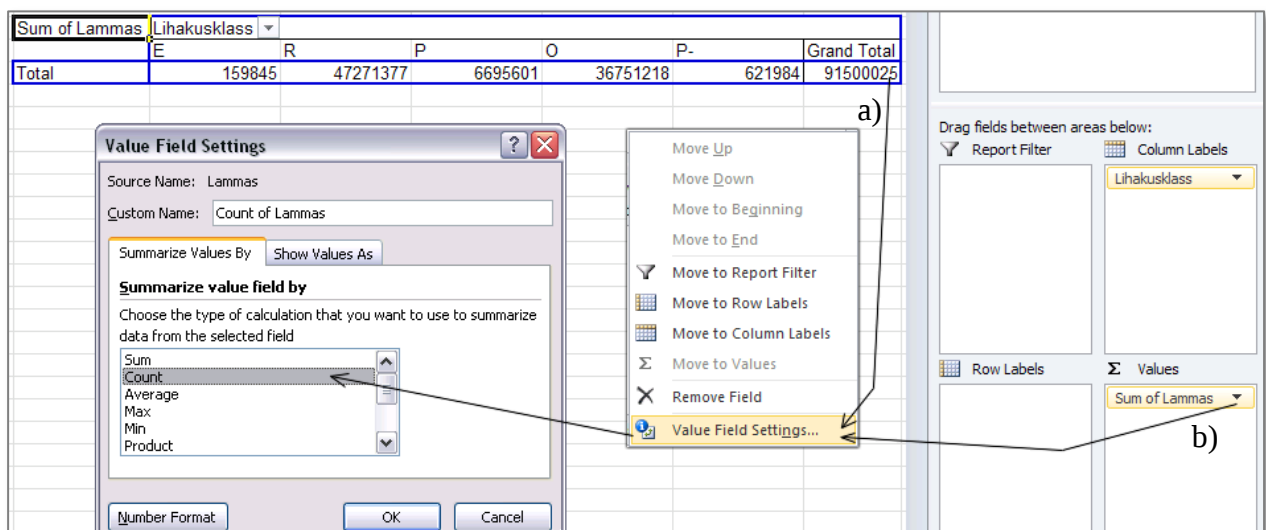
- Vaikimisi arvutab Excel lammaste numbrite summa ...
Et selle asemel lihtsalt kokku lugeda, kui mitu lammast mingisse lihakusklassi kuulus, tuleb ära muuta *PivotTable*-s kasutatav funktsioon (*Sum* asemel *Count*):

PivotTable Tools-sakk → *Options* → *Summarize Values By*



Alternatiivina võib *PivotTable*-s rakendatavat funktsiooni muuta ka

- klikkides tabelil hiire parempoolse nupuga või
- klikkides *PivotTable* 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 (klõpsates selleks klassi nimel (näiteks lahtril 'P') ja tõstes lahtri servast kinni hoides õigesse kohta):

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

Tulemus:

Count of Lammas	Lihakusklass					
	E	R	O	P	P-	Grand Total
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*);

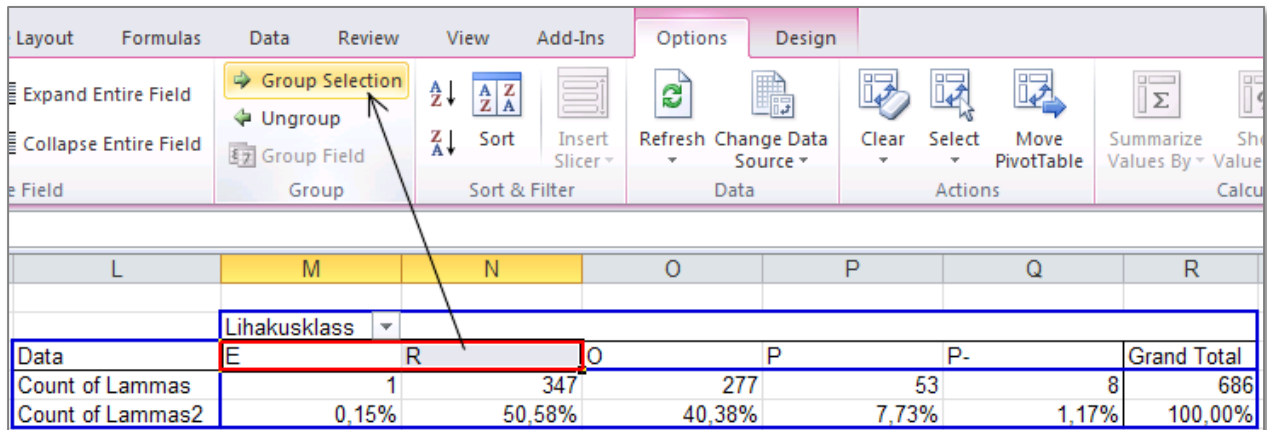
c) määrake vajadusel ka uue rea tarvis funktsiooniks *Count* (*Sum* asemel) ning

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

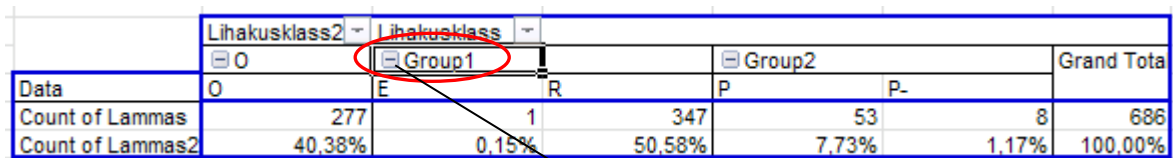
Tulemus:

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

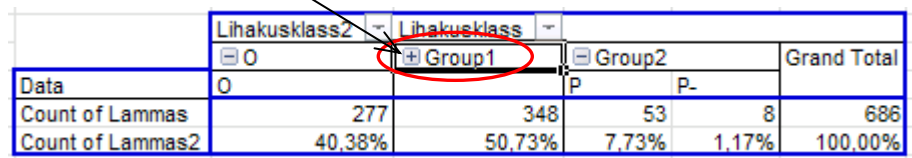
- Et paremaks kui „hea“ (kood „E“) on hinnatud vaid üht rümpa, võiks selle ühendada grupiga „R“ (moodustada uus grupp – vähemalt hindega „hea“ rümpad).



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

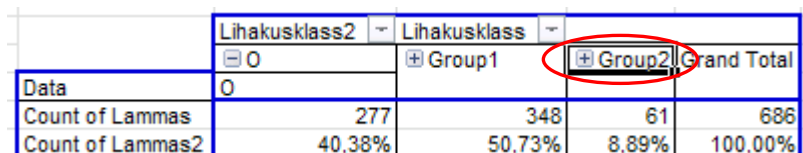


	Lihakusklass2	Lihakusklass					
	O	Group1	Group2				Grand Total
Data	O	E	R	P	P-		
Count of Lammas	277	1	347	53	8		686
Count of Lammas2	40,38%	0,15%	50,58%	7,73%	1,17%		100,00%



	Lihakusklass2	Lihakusklass					
	O	Group1	Group2				Grand Total
Data	O		P	P-			
Count of Lammas	277	348	53	8			686
Count of Lammas2	40,38%	50,73%	7,73%	1,17%			100,00%

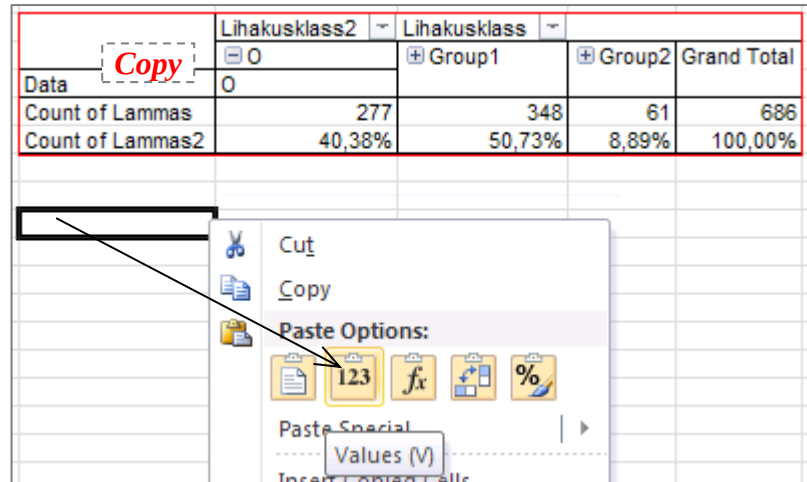
Grupeerige analoogsel viisil ka „lahjad“ ja „eriti lahjad“ rümpad (grupid „P“ ja „P-“, sest ega seal suurt vahet pole).



	Lihakusklass2	Lihakusklass					
	O	Group1	Group2				Grand Total
Data	O						
Count of Lammas	277	348	61				686
Count of Lammas2	40,38%	50,73%	8,89%				100,00%

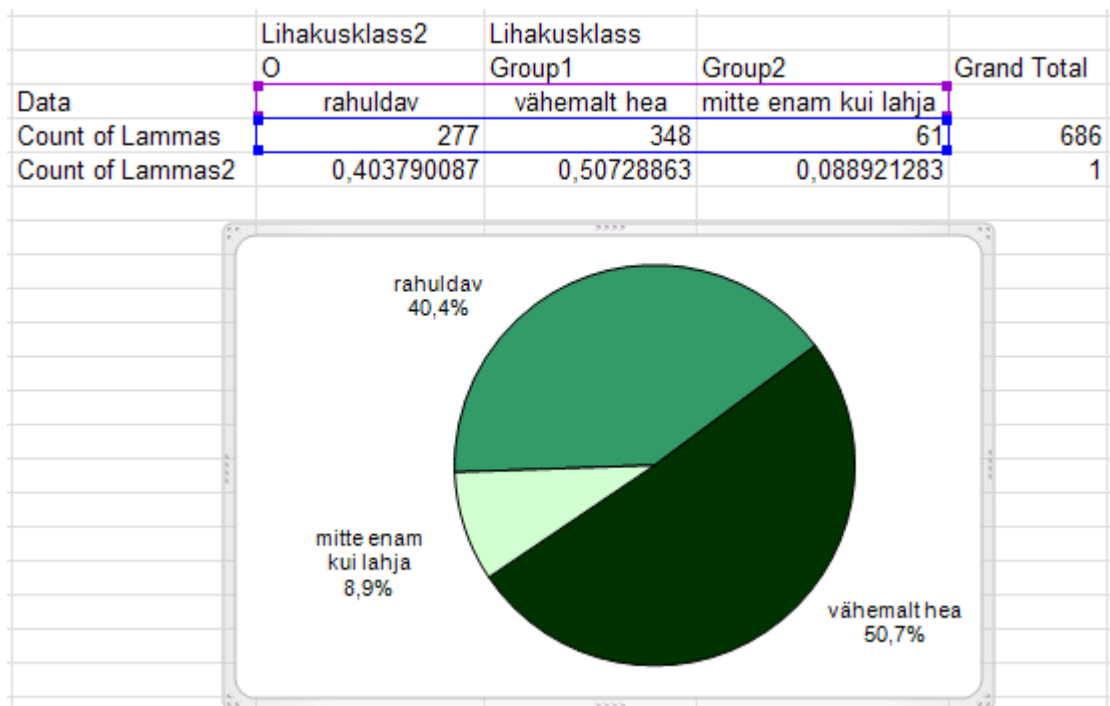
2. *PivotTable*'i abil konstrueeritud tabeli põhjal kenade jooniste tegemiseks on sageli mõttekas teha vajalikest väärtustest abitabel ja joonistada diagramm abitabeli alusel.

Põhjuseks on see, et otse *PivotTable* alusel joonise tegemise tulemuseks on nn *PivotChart*, mis on sarnaselt *PivotTable*'ga lingitud andmetabeliga, seeläbi kergesti täiendatav ja ümberarvutatav, aga ei võimalda muuta kõike tavalisel Exceli 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 esinemis-sagedused, muutke soovi korral värve):



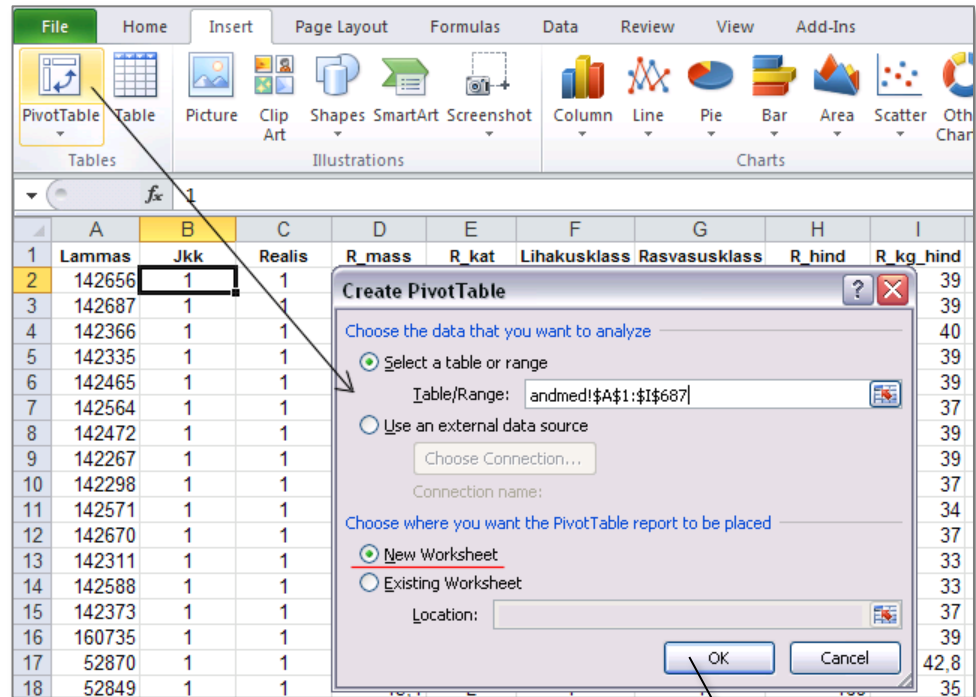
3. Viimaks konstrueerige sagedustabelid ja joonistage nende alusel sektordiagrammid ka rümba (üld)kategoriale ja rasvasusklassile (kui tundub, et aega üle jääb).

ÜLESANNE 2.

--- Kahemõõtmeline sagedustabel ja χ^2 -test ---

Kas rümpade jagunemine rasvasusklassidesse sõltub rümba üldkategorias?

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



Drop Report Filter Fields Here						
		Rasvasusklass				
R kat	Data	1	2	3	4	5 Grand Total
L	Sum of Lammas	9224926	47487107	20998406	2184105	1801598 81696142
	Sum of Lammas2	9224926	47487107	20998406	2184105	1801598 81696142
	Sum of Lammas3	9224926	47487107	20998406	2184105	1801598 81696142
S	Sum of Lammas	251825	3447190	2319642	1341062	2444164 9803883
	Sum of Lammas2	251825	3447190	2319642	1341062	2444164 9803883
	Sum of Lammas3	251825	3447190	2319642	1341062	2444164 9803883
Total Sum of Lammas		9476751	50934297	23318048	3525167	4245762 91500025
Total Sum of Lammas2		9476751	50934297	23318048	3525167	4245762 91500025
Total Sum of Lammas3		9476751	50934297	23318048	3525167	4245762 91500025

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

R kat	Data					
L	Sum of Lammas					
	Sum of Lammas2					
	Sum of Lammas3					
S	Sum of Lammas					
	Sum of Lammas2					
	Sum of Lammas3					
Total	Sum of Lammas					
Total	Sum of Lammas2					
Total	Sum of Lammas3					

R kat	Data	Rasvasusklass	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 teine ja kolmas vaatluste arv (*Count of Lammas 2 ja 3*) vastavalt rea- ja veeru-
protsendina:

R kat	Data	Rasvasusklass	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	Rasvasusklass	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)

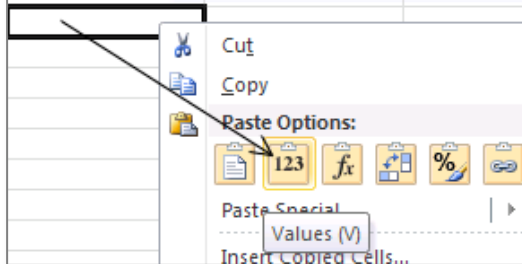
2. Testige rümpade üldkategoriatesse ja rasvasuklassidesse jagunemise vahelise seose statistilist olulisust.

- Esmalt – et oleks selge, mida te üldse testite – pange esmalt kirja kontrollitav hüpoteeside paar).
- Järgnevalt konstrueerige uus kahemõõtmeline sagedustabel, mis sisaldab üksnes absoluutseid sagedusi,

H_0 (nullhüpotees): ...						
H_1 (alternatiivne hüpotees): ...						
Empiirilised (andmetest arvutatud) sagedused						
Count of Lammas	Rasvasusklass					
R_kat	1	2	3	4	5	Grand Total
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 arvutatud) sagedused						
Count of Lammas	Rasvasusklass					
R_kat	1	2	3	4	5	Grand Total
L	69	301	132	14	10	526
S	2	35	58	34	31	160
Grand Total	71	336	190	48	41	686



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

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

- Arvutage uude tabelisse kustutatud sageduste asemele tunnuste sõltumatus juhule (nullhüpoteesile) vastavad nõ teoreetilised sagedused
(kujul: reasumma $\times$ veerusumma / vaatluste arv).

	A	B	C	D	E	F	G
25							
26	Teoreetilised (nullhüpoteesile e sõltumatus 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 sõltumatus 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 (sõltumatus juhule vastavate) sagedustega ja väljastab olulisuse tõenäosuse p väärtuse

Et oleks lihtsam aru saada, milliste arvude võrdlemisel χ^2 -test baseerub (ehk siis millised tabelite osad tuleb Exceli 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).

=CHISQ.TEST(B21:F22;B29:F30)							
	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 sõltumatus 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 lõppjärelus (Kas seos on statistiliselt oluline? Miks te nii otsustasite?).

Hii-ruut-test 3,69291E-37

ÜLESANNE 3.

--- Regressioonianalüüs graafiliselt ---

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

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

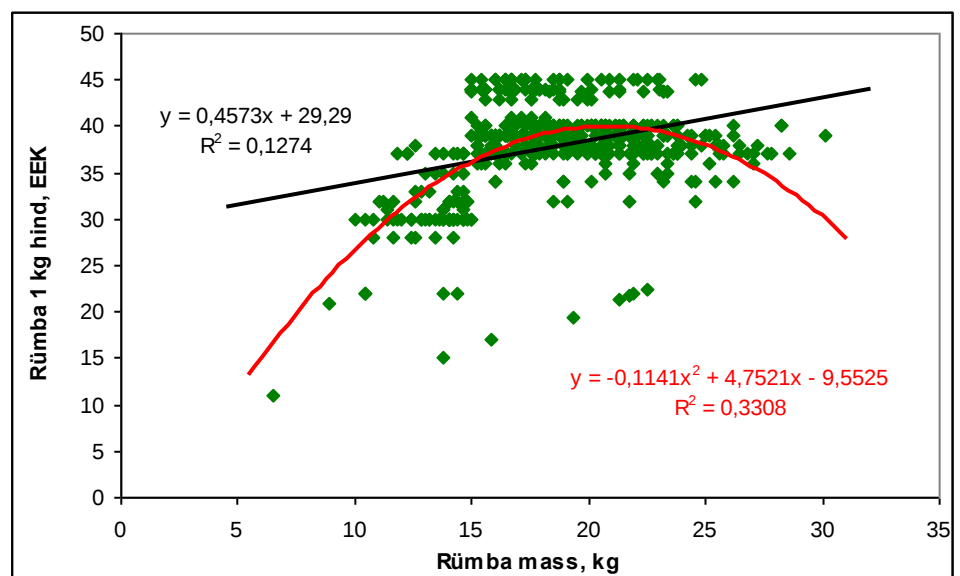
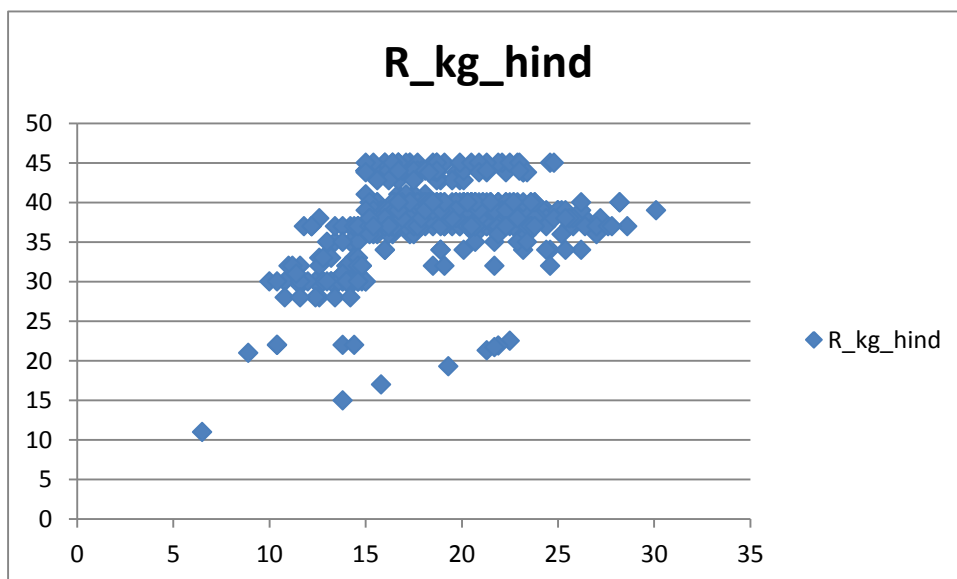
The screenshot shows the Excel interface with the 'Sort & Filter' dropdown menu open for column E (R_kat). The menu options include 'Sort A to Z', 'Sort Z to A', 'Custom Sort...', 'Filter', and 'Clear'. The 'Filter' option is highlighted. Below the menu, a detailed filter view for column E is shown, displaying various filtering options like 'Sort A to Z', 'Sort Z to A', 'Sort by Color', 'Clear Filter From "R_kat"', 'Filter by Color', 'Text Filters', and a search box. The 'Filter' button is also highlighted in this view.

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	Lihakuskla	Rasvasuskla	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

2. 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 punktide parvest läbi ka ruut-funktsiooni graafik ning tellige sellegi tarvis Excelilt 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?



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