

## Praks 2

### Eel- ja järeltöö

1. Salvestage arvutisse andmestik lammas.xls ([http://ph.emu.ee/~ktanel/VL\\_1112/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](mailto: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  $\chi^2$ -testi ja regressioonanalüüsi.

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 lihaku- 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 (ueele 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  $\chi^2$ -test – viimane võrdleb empiirilisi (andmetabelist arvatud) sagedusi teoreetiliste (sõltumatuse juhule vastavate) sagedustega ja väljastab olulisuse tõenäosuse  $p$  väärtuse.
  - Sõnastage lõppjärelus (viidates sõnastuses ka  $p$ -väärtusele, millel järeldus 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 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 punktiparvest 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.
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).
  - Sorteeri (või filtreeri) 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 lõppjärelendus.

**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.

*Data / Andmed* → *PivotTable and PivotChart Report...* / *PivotTable ja PivotChart Aruanne...*

The screenshot illustrates the process of creating a PivotTable in Microsoft Excel. The background shows a spreadsheet with columns A through I and rows 1 through 5. The data in the spreadsheet is as follows:

	A	B	C	D	E	F	G	H	I
1	Lammas	Jkk	Realis	R_mass	R_kat				R_kg_hind
2	142656	1	1	23,8	L				39
3	142687	1	1	15,2	L				39
4	142366	1	1	15,6	L	R	2		624
5	142335	1	1	24,4					40

The PivotTable and PivotChart Wizard is shown in three steps:

- Step 1 of 3:** "Where is the data that you want to analyze?" The "Microsoft Office Excel list or database" option is selected. "What kind of report do you want to create?" The "PivotTable" option is selected.
- Step 2 of 3:** "Where is the data that you want to use?" The Range is set to "\$A\$1:\$I\$687".
- Step 3 of 3:** "Where do you want to put the PivotTable report?" The "Existing worksheet" option is selected, and the location is set to "andmed!\$M\$2". A red arrow points to cell M2 in the spreadsheet, with the text "Loodava tabeli vasaku ülemise nurga asukoht" (Location of the top-left corner of the new table) next to it.

Drop Column Fields Here

Drop Row Fields Here

Drop Data Items Here

**PivotTable Field List**

Drag items to the PivotTable report

- Lammas
- Jkk
- Realis
- R\_mass
- R\_kat
- Lihakusklass
- Rasvasusklass
- R\_hind

Add To Row Area

	E	R	P	O	P-	Grand Total
Total		1	347	53	277	686

**PivotTable Field**

Source field: Lammas

Name: Count of Lammas

Summarize by:

- Sum
- Count
- Average
- Max
- Min
- Product
- Count Nums

OK Cancel Hide Number... Options >>

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

Sorteeri lihakusklassid sisulise järjestuse alusel (*Excel* ei pruugi seda alati õigesti teha) – lihtsaim võimalus vaid mõned üksikut gruppi ümber tõsta on klikkida selle nimel (näiteks lahtril 'P') ja tõsta see lahtri servast kinni hoides õigesse kohta.

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

Q3:Q4

Lisaks absoluutsetele sagedustele suhteliste sageduste arvutamine.

The image illustrates the process of calculating relative frequencies from absolute counts in a PivotTable. It shows the PivotTable Field List, the PivotTable Field dialog box, and the resulting PivotTable with relative frequencies.

**PivotTable Field List:**

- Source: Lammas
- Fields: Jkk, Realis, R\_mass, R\_kat, **Lihakusklass**, Rasvasusklass, R\_hind

**PivotTable Field Dialog (Initial):**

- Source field: Lammas
- Name: Count of Lammas2
- Summarize by: Count

**PivotTable Field Dialog (Final):**

- Source field: Lammas
- Name: Count of Lammas2
- Summarize by: Count
- Show data as: % of column

**PivotTable Data:**

	E	R	O	P	P-	Grand Total
Count of Lammas	1	347	277	53	8	686
Sum of Lammas	159845	47271377	36751218	6695601	621984	91500025

**Final PivotTable (Relative Frequencies):**

	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 „R“) on hinnatud vaid üht rümpa, võiks selle kas edasisest analüüsist välja jätta

	O	P	P-	Grand Total
Count of Lammas	277	53	8	686
Count of Lammas2	40,38%	7,73%	1,17%	100,00%

või ühendada grupiga „R“ (moodustada uus grupp – vähemalt hindega „hea“ rümbad).

	O	P	P-	Grand Total
Count of Lammas	277	53	8	686
Count of Lammas2	40,38%	7,73%	1,17%	100,00%

	O	P	P-	Grand Total	
Count of Lammas	347	277	53	8	686
Count of Lammas2	58%	40,38%	7,73%	1,17%	100,00%

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

Grupeerige analoogsel viisil ka „lahjad“ ja „eriti lahjad“ rümbad (sest ega seal suurt vahet pole).

	Lihakusklass2	Lihakusklass		
Data	Group1	0	Group2	Grand Total
Count of Lammas	348	277	61	686
Count of Lammas2	50,73%	40,38%	8,89%	100,00%

*Pivot Table*'i abil konstrueeritud tabeli põhjal kenade jooniste tegemiseks on sageli soovitatav teha vajalikest väärtustest abitabel ja teha joonis selle 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).

The screenshot illustrates the workflow for creating a chart from a PivotTable. It shows the PivotTable with the 'Count of Lammas' data. A context menu is open over the table, and the 'Copy' option is selected. The 'Paste Special' dialog box is shown with 'Values' selected under the 'Paste' section. A small table is extracted from the PivotTable, and a pie chart is generated based on its data.

vähemalt hea	rahuldav	mitte enam kui lahja
348	277	61

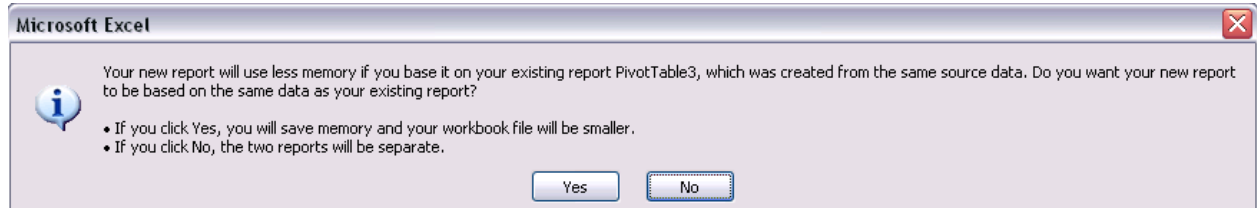
The pie chart shows the following distribution:

- vähemalt hea: 50,7%
- rahuldav: 40,4%
- mitte enam kui lahja: 8,9%



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

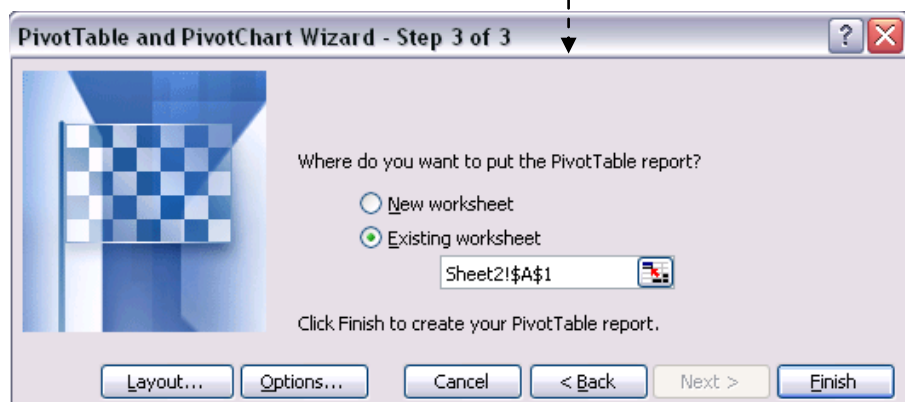
NB! Uue *PivotTable*'i tegemisel ilmuvale lisaküsimusele on kõige lollikindlam vastata 'No':



## 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.

	A	B	C	D	E					
1	Lammas	Jkk	Realis	R_mass	R_kat					
2	142656	1	1	23,8	L					
3	142687	1	1	15,2	L					
4	142366	1	1	15,6	L	R	2		624	40
5	142335	1	1	24,4	L	R	3		951,6	39



Drop Column Fields Here

Drop Row Fields Here

Drop Data Items Here

**PivotTable Field List**

Drag items to the PivotTable report

- Lammas
- Jkk
- Realis
- R\_mass
- R\_kat
- Lihakusklass
- Rasvasusklass
- R\_hind

Add To Row Area

**PivotTable Field**

Source field: Lammas

Name: Count of Lammas2\_2

Summarize by:

- Sum
- Count
- Average
- Max
- Min
- Product
- Count Nums

Show data as:

- % of column
- Normal
- Difference From
- % Of
- % Difference From
- Running Total in
- % of row
- % of column

LihakusKlass

R_kat	Data	Rasvasusklass					Grand Total
		1	2	3	4	5	
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 Lammas2_2	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 Lammas2_2	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 Lammas2_2		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Kommentaariid. ....

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

H<sub>0</sub> (nullhüpotees): ...  
 H<sub>1</sub> (alternatiivne hüpotees): ...

**Empiirilised (andmetest arvatud) sagedused**

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

**Teoreetilised (nullhüpoteesile e sõltumatus juhule vastavad) sagedused**

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

**Paste Special**

Paste

- All
- Formulas
- Values
- Formats
- Comments
- Validation
- All except borders
- Column widths
- Formulas and number formats
- Values and number formats

Operation

- None
- Add
- Subtract
- Multiply
- Divide

Skip blanks     Transpose

Paste Link    OK    Cancel

Teoreetiliste sageduste arvutamine:

	A	B	C	D	E	F	G
25							
26	<b>Teoreetilised (nullhüpoteesile e sõltumatus juhule vastavad) sagedused</b>						
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	<b>Teoreetilised (nullhüpoteesile e sõltumatus juhule vastavad) sagedused</b>						
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

$\chi^2$ -testi saab *Excel*'s teostada funktsiooniga CHISQ.TEST (*Excel* 2003-s CHITEST; tulemuseks on olulisuse tõenäosus  $p$ ).

Et oleks lihtsam aru saada, milliste arvude võrdlemisel  $\chi^2$ -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).

Empiirilised (andmetest arvatatud) 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

Teoreetilised (nullhüpooteesile e sõltumatus juhule vastavad) sagedused						
Count of Lammas	Rasvasusklass					
R_kat	1	2	3	4	5	Grand Total
L	54,44023324	257,6326531	145,6851312	36,80466472	31,43731778	526
S	16,55976676	78,36734694	44,3148688	11,19533528	9,562682216	160
Grand Total	71	336	190	48	41	686

Hii-ruut-test

=

CHISQ.TEST

**Function Arguments**

CHITEST

Actual\_range B21:F22 = {69;301;132;14;10;}

Expected\_range B29:F30 = {54,4402332361516

= 3,69291E-37

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Expected\_range is the range of data that contains the ratio of the product of row totals and column totals to the grand total.

Formula result = 3,69291E-37

[Help on this function](#)

OK Cancel

**Insert Function**

Search for a function:

Type a brief description of what you want to do and then click Go

Or select a category: Statistical

Select a function:

CHIDIST  
CHIINV  
**CHITEST**  
CONFIDENCE  
CORREL  
COUNT  
COUNTA

**CHITEST(actual\_range;expected\_range)**

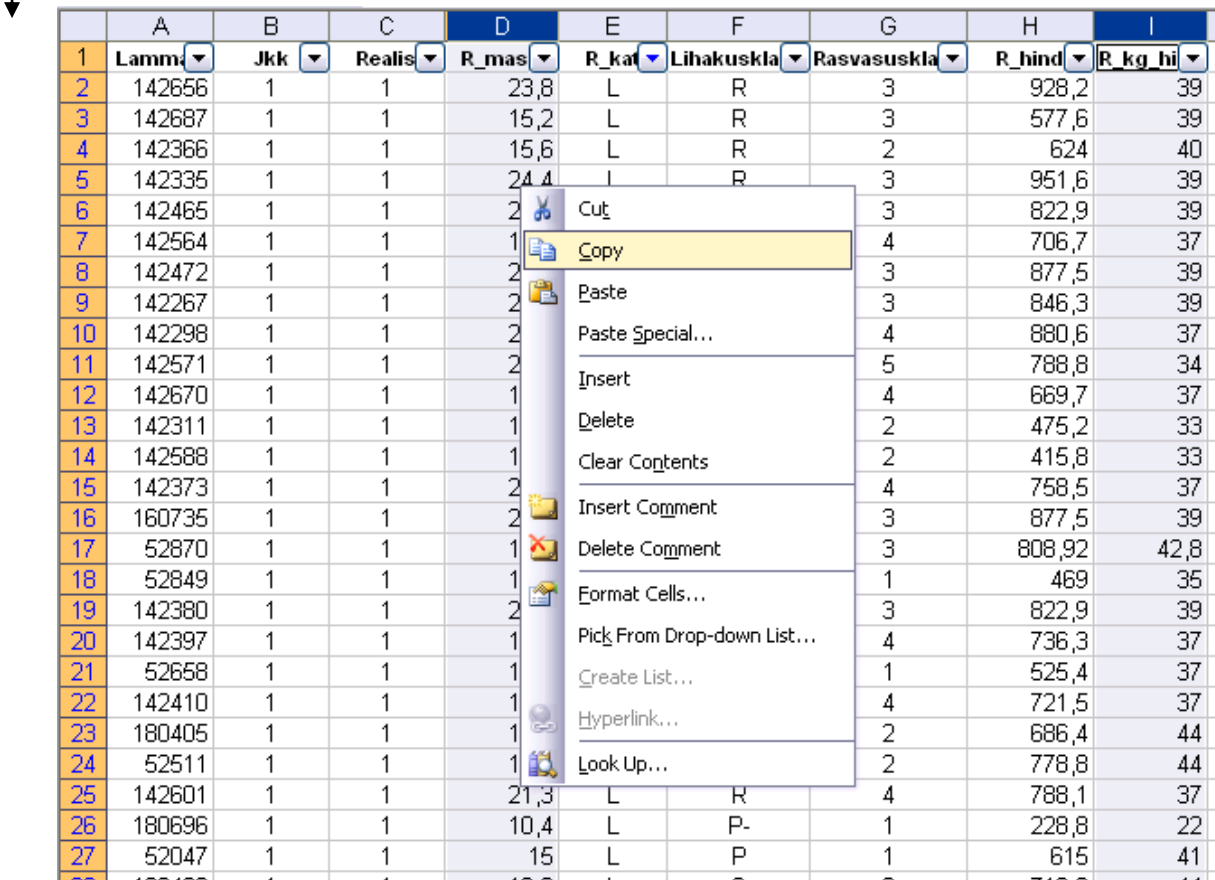
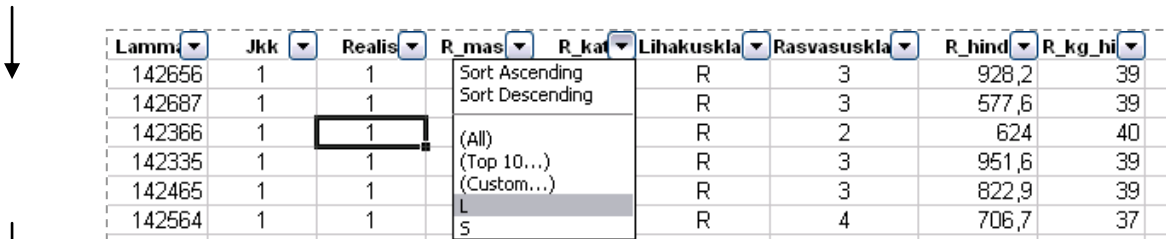
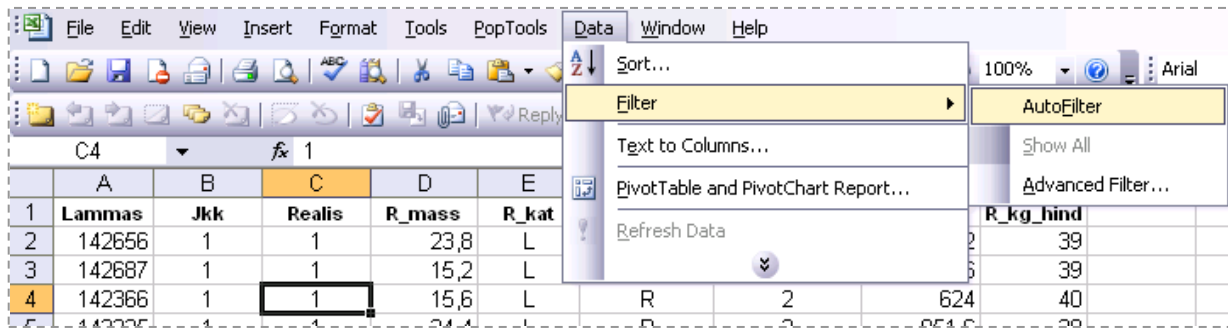
Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

[Help on this function](#)

OK Cancel

Sõnastage lõppjärelendus (Kas seos on statistiliselt oluline? Miks te nii otsustasite?).

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



	A	B	C	D	E	F	G	H	I
1	R_mass	R_kg_hind							
2	23,8	39							
3	15,2	39							
4	15,6	40							
5	24,4	39							
6	21,1	39							
7	19,1	37							
8	22,5	39							
9	21,7	39							
10	23,8	37							
11	23,2	34							
12	18,1	37							
13	14,4	33							
14	12,6	33							
15	20,5	37							
16	22,5	39							
17	18,9	42,8							
18	13,4	35							
19	21,1	39							
20	19,9	37							
21	14,2	37							
22	19,5	37							
23	15,6	44							
24	17,7	44							
25	21,3	37							
26	10,4	22							
27	15	41							
28	16,2	44							
29	26,2	34							

