

Biomeetria praks 7

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excel*'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks7' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks7' ülemisse vasakusse nurka.

Ülesanne.

Kas gümnaasiumi matemaatika hinne ja auto omamine on seotud?

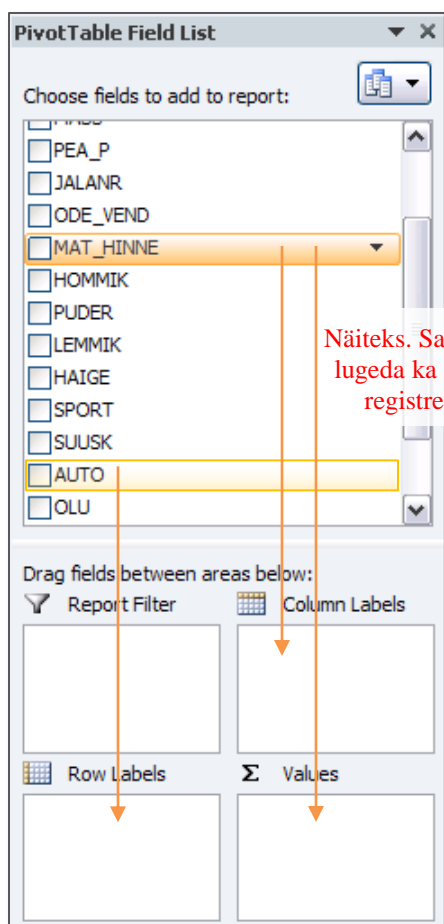
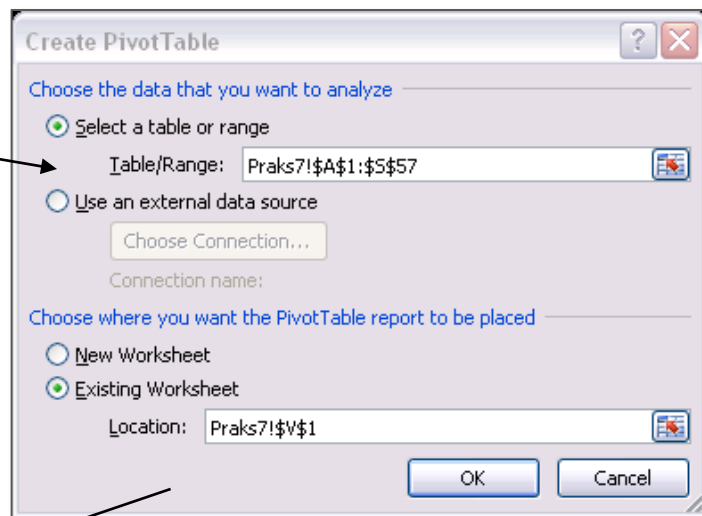
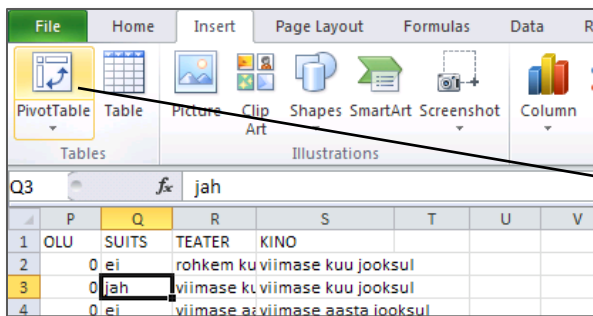
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele ' MAT_HINNE ' ja 'AUTO'.
2. Leidke sellesse tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii veeru- kui ka reaprotsendid).
3. Kommenteerige tabelit (sõnastades laused vähemalt kahe arvutatud protsendi kohta)!
4. Testimaks tuvastatud seose/erinevuse statistilist olulisust, pange kirja vastav hüpoteeside paar.
5. Tehke uus 2-mõõtmeline sagedustabel tunnustele ' MAT_HINNE ' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi;
6. arvutage viimase alusel nullhüpoteesile (seost ei ole) vastavad sagedused ja
7. viige läbi hii-ruut-test (leidke p -väärtus).
8. Sõnastage lõppjärelendus (kas seos on statistiliselt oluline, mille alusel te seda otsustasite ja milles see seos ikkagi seisneb?).

Tööjuhend

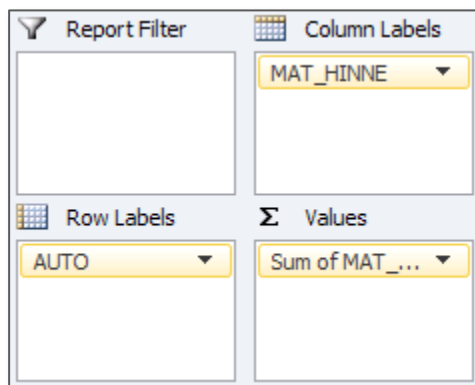
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'MAT_HINNE' ja 'AUTO'.

- Paigutage kursor andmetabeli suvalisse lahtrisse (leheküljel 'Praks 7').
- *Insert*-sakk → *PivotTable*

Loodav tabel paigutage samale 'Praks7'-lehele:

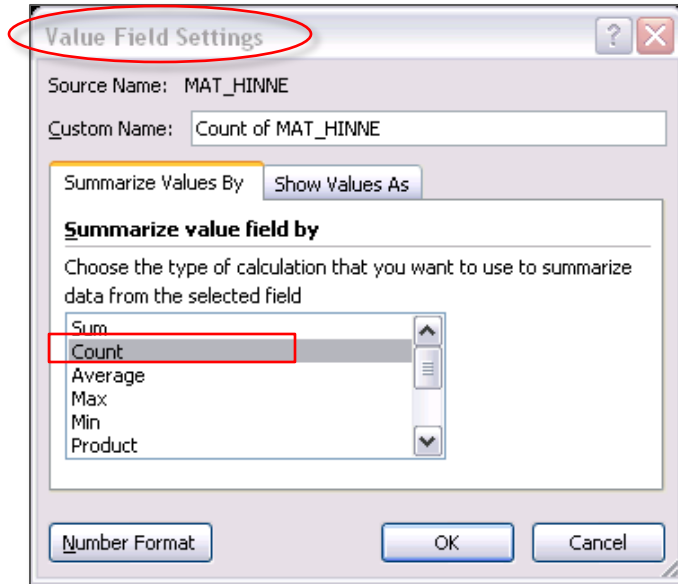


Näiteks. Samas võib lasta *Excelil* kokku lugeda ka mõne teise **kõigil** tudengitel registreeritud tunnuse väärtused.

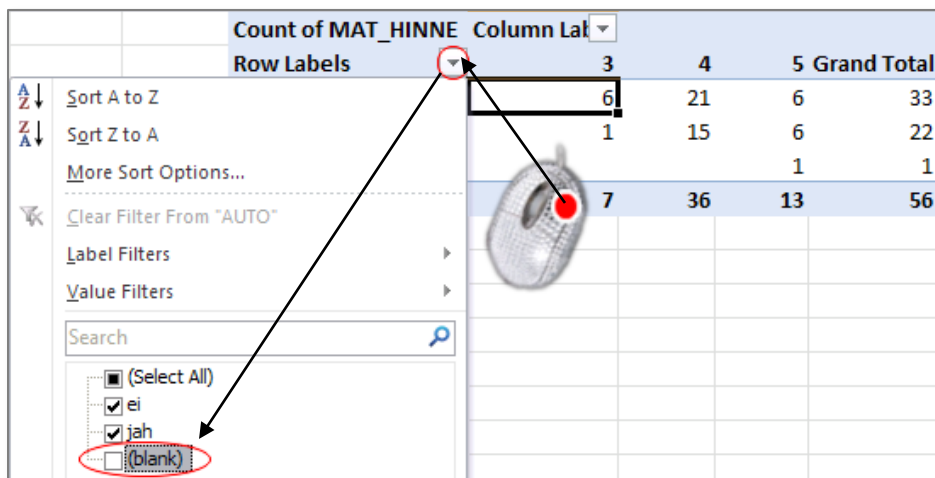


| Sum of MAT_HINNE | Column Labels | 3 | 4 | 5 | Grand Total |
|--------------------|---------------|-----------|------------|-----------|-------------|
| Row Labels | | | | | |
| ei | | 18 | 84 | 30 | 132 |
| jah | | 3 | 60 | 30 | 93 |
| (blank) | | | | 5 | 5 |
| Grand Total | | 21 | 144 | 65 | 230 |

- Arvutage tabelisse summa asemel tudengite arv.



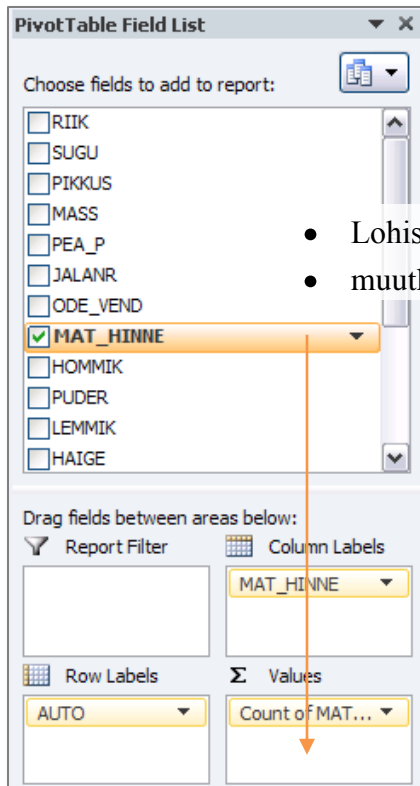
- Kaotage tabelist puuduvate väärtuste (blank) vastav rida.



Tulemus:

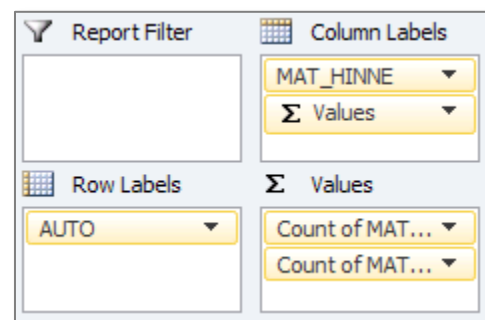
| Count of MAT_HINNE | 3 | 4 | 5 | Grand Total |
|--------------------|----------|-----------|-----------|-------------|
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |

2. Leidke tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii rea- kui ka veeruprotsendid).



- Lohistage tunnus 'MAT_HINNE' ka teine kord väärtuste lahtrisse ja
- muutke funktsioon 'Summa / Sum' funktsiooniks 'Loendus / Count'

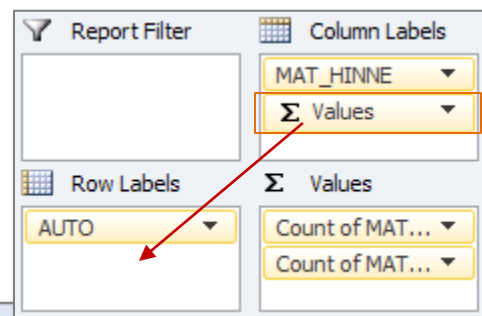
Tulemus:



| Column Labels | 3 | 4 | 5 | Total Count of MAT_HINNE | Total Count of MAT_HINNE2 |
|---------------|--------------------|---------------------|--------------------|--------------------------|---------------------------|
| Row Labels | Count of MAT_HINNE | Count of MAT_HINNE2 | Count of MAT_HINNE | Count of MAT_HINNE2 | Count of MAT_HINNE2 |
| ei | 6 | 6 | 21 | 21 | 6 |
| jah | 1 | 1 | 15 | 15 | 6 |
| Grand Total | 7 | 7 | 36 | 36 | 12 |

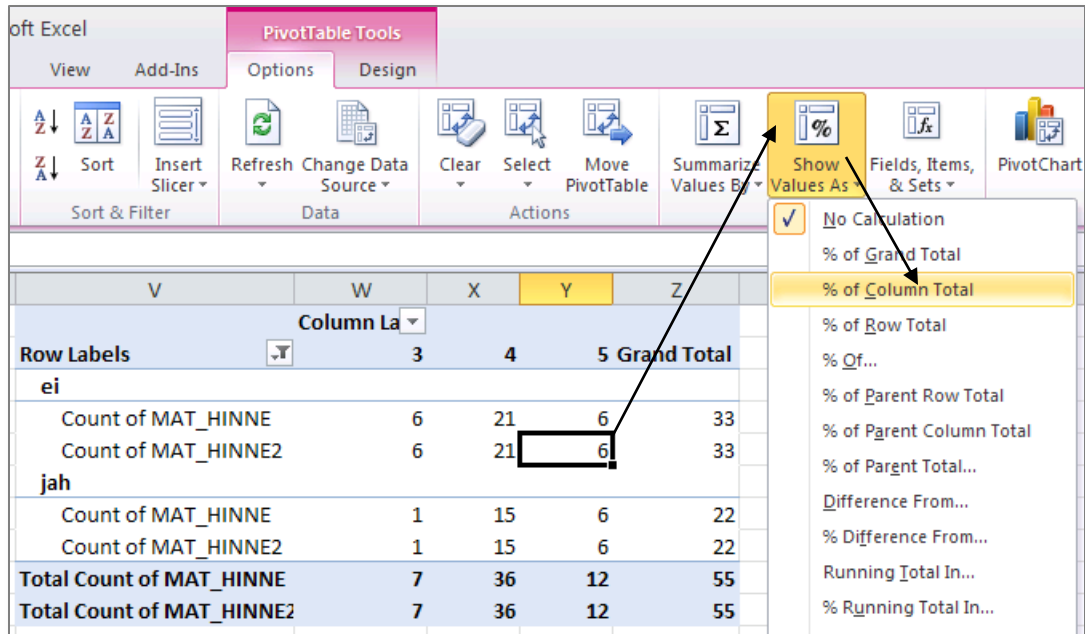
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu Excel vaikimisi teeb):

Tulemus:



| Row Labels | 3 | 4 | 5 | Grand Total |
|----------------------------------|----------|-----------|-----------|-------------|
| ei | | | | |
| Count of MAT_HINNE | 6 | 21 | 6 | 33 |
| Count of MAT_HINNE2 | 6 | 21 | 6 | 33 |
| jah | | | | |
| Count of MAT_HINNE | 1 | 15 | 6 | 22 |
| Count of MAT_HINNE2 | 1 | 15 | 6 | 22 |
| Total Count of MAT_HINNE | 7 | 36 | 12 | 55 |
| Total Count of MAT_HINNE2 | 7 | 36 | 12 | 55 |

- Esitamaks *Exceli* poolt uuesti kokku loetud tudengite arvu veeruprotsendina:



Tulemus:

| Row Labels | 3 | 4 | 5 | Grand Total |
|----------------------------------|----------------|----------------|----------------|----------------|
| ei | | | | |
| Count of MAT_HINNE | 6 | 21 | 6 | 33 |
| Count of MAT_HINNE2 | 85,71% | 58,33% | 50,00% | 60,00% |
| jah | | | | |
| Count of MAT_HINNE | 1 | 15 | 6 | 22 |
| Count of MAT_HINNE2 | 14,29% | 41,67% | 50,00% | 40,00% |
| Total Count of MAT_HINNE | 7 | 36 | 12 | 55 |
| Total Count of MAT_HINNE2 | 100,00% | 100,00% | 100,00% | 100,00% |

- Arvutage analoogselt ka reaprotsendid ... (% of Row Total).

(alustage sarnaselt punktis 2 kirjeldatule)

| Row Labels | 3 | 4 | 5 | Grand Total |
|----------------------------------|----------------|----------------|----------------|----------------|
| ei | | | | |
| Count of MAT_HINNE | 6 | 21 | 6 | 33 |
| Count of MAT_HINNE2 | 85,71% | 58,33% | 50,00% | 60,00% |
| Count of MAT_HINNE3 | 18,18% | 63,64% | 18,18% | 100,00% |
| jah | | | | |
| Count of MAT_HINNE | 1 | 15 | 6 | 22 |
| Count of MAT_HINNE2 | 14,29% | 41,67% | 50,00% | 40,00% |
| Count of MAT_HINNE3 | 4,55% | 68,18% | 27,27% | 100,00% |
| Total Count of MAT_HINNE | 7 | 36 | 12 | 55 |
| Total Count of MAT_HINNE2 | 100,00% | 100,00% | 100,00% | 100,00% |
| Total Count of MAT_HINNE3 | 12,73% | 65,45% | 21,82% | 100,00% |

3. Kommenteerige tabelit (sõnastades laused vähemalt ühe rea- ja ühe veeruprotsendi kohta)!

4. Seose statistilise olulisuse testimiseks **pange** esmalt **kirja hüpoteeside paar** (et oleks ikka selge, mida testite).

5. Tehke uus 2-mõõtmeline sagedustabel tunnustele 'MAT_HINNE' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi.

| Row Labels | 3 | 4 | 5 | Grand Total |
|---|----------------|----------------|----------------|----------------|
| ei | | | | |
| Count of MAT_HIN | 6 | 21 | 6 | 33 |
| Count of MAT_HIN | 85,71% | 58,33% | 50,00% | 60,00% |
| Count of MAT_HIN | 18,18% | 63,64% | 18,18% | 100,00% |
| jah | | | | |
| Count of MAT_HIN | 1 | 15 | 6 | 22 |
| Count of MAT_HIN | 14,29% | 41,67% | 50,00% | 40,00% |
| Count of MAT_HIN | 4,55% | 68,18% | 27,27% | 100,00% |
| Total Count of MAT_HI | 7 | 36 | 12 | 55 |
| Total Count of MAT_HI | 100,00% | 100,00% | 100,00% | 100,00% |
| Total Count of MAT_HI | 12,73% | 65,45% | 21,82% | 100,00% |
| Kommentaarid. | | | | |
| H ₀ : Matemaatika hinne ja auto omamine ei ole seotud. | | | | |
| H ₁ : Matemaatika hinne ja auto omamine on seotud. | | | | |
| Count of MAT_HINNE | 3 | 4 | 5 | Grand Total |
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |

6. Tehke uus, viimati loodud tabeliga analoogne sagedustabel, kuhu arvutage nullhüpoteesile vastavad sagedused.

Kuidas? Vt järgnevaid alapunkte!

- Tehke koopia *Pivot Table*'i abil konstrueeritud tabeli struktuurist ja väärtustest:

| Count of MAT_HINNE | 3 | 4 | 5 | Grand Total |
|--------------------|----------|-----------|-----------|-------------|
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |

| Count of MAT_HINNE | 3 | 4 | 5 | Grand Total |
|--------------------|----------|-----------|-----------|-------------|
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |

- seejärel kustutage ära kopeeritud tabeli **sisu**

| Count of MAT_HINNE | 3 | 4 | 5 | Grand Total |
|--------------------|--------------|---------------|--------------|-------------|
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |

- ning arvutage sinna asemele sõltumatuse juhule (nullhüpooteesile) vastavad sagedused (valemist $n_{ij} = n_i \times n_j / n$).

| | V | W | X | Y | Z |
|----|----------------------------------|--------------|----|----|-------------|
| 29 | Count of MAT_HINNE Column Labels | | | | |
| 30 | Row Labels | 3 | 4 | 5 | Grand Total |
| 31 | ei | =Z31*W33/Z33 | | | 33 |
| 32 | jah | | | | 22 |
| 33 | Grand Total | 7 | 36 | 12 | 55 |

$n_{1.}$
.....

| | V | W | X | Y | Z |
|----|----------------------------------|-----|------|-----|-------------|
| 29 | Count of MAT_HINNE Column Labels | | | | |
| 30 | Row Labels | 3 | 4 | 5 | Grand Total |
| 31 | ei | 4,2 | 21,6 | 7,2 | 33 |
| 32 | jah | 2,8 | 14,4 | 4,8 | 22 |
| 33 | Grand Total | 7 | 36 | 12 | 55 |

$n_{.2}$

7. Viige läbi hii-ruut test (leidke p -väärtus) – funktsioon CHISQ.TEST.

Et oleks lihtsam aru saada, milliste arvude võrdlemisel χ^2 -test baseerub (ehk siis millised tabelite osad tuleb Excel'i funktsioonile CHISQ.TEST ette anda), võib vastavad lahtrid selguse mõttes näiteks ära värvida ja/või kirjutada tabelitele peale, milliste sagedustega tegu on.

| U | V | W | X | Y | Z |
|----|---|-----|------|-----|-------------|
| 20 | | | | | |
| 21 | Tegelikud andmete alusel arvutatud (empiirilised) sagedused | | | | |
| 22 | Count of MAT_HINNE Column Labels | | | | |
| 23 | Row Labels | 3 | 4 | 5 | Grand Total |
| 24 | ei | 6 | 21 | 6 | 33 |
| 25 | jah | 1 | 15 | 6 | 22 |
| 26 | Grand Total | 7 | 36 | 12 | 55 |
| 27 | | | | | |
| 28 | Teoreetilised nullhüpooteesile vastavad sagedused | | | | |
| 29 | Count of MAT_HINNE Column Labels | | | | |
| 30 | Row Labels | 3 | 4 | 5 | Grand Total |
| 31 | ei | 4,2 | 21,6 | 7,2 | 33 |
| 32 | jah | 2,8 | 14,4 | 4,8 | 22 |
| 33 | Grand Total | 7 | 36 | 12 | 55 |
| 34 | | | | | |
| 35 | | | | | |
| 36 | Hii-ruut-test | | | | |

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

NB! Vanemais Exceli versioonides testib sama hüpooteesi funktsioon CHITEST.

Tulemus:

| | | | | |
|---|------------|------|-----|-------------|
| H ₀ : Matemaatika hinne ja auto omamine ei ole seotud. | | | | |
| H ₁ : Matemaatika hinne ja auto omamine on seotud. | | | | |
| <u>Tegelikud andmete alusel arvatud (empiirilised) sagedused</u> | | | | |
| Count of MAT_HINNE Column Labels | | | | |
| Row Labels | 3 | 4 | 5 | Grand Total |
| ei | 6 | 21 | 6 | 33 |
| jah | 1 | 15 | 6 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |
| <u>Teoreetilised nullhüpoteesile vastavad sagedused</u> | | | | |
| Count of MAT_HINNE Column Labels | | | | |
| Row Labels | 3 | 4 | 5 | Grand Total |
| ei | 4,2 | 21,6 | 7,2 | 33 |
| jah | 2,8 | 14,4 | 4,8 | 22 |
| Grand Total | 7 | 36 | 12 | 55 |
| Hii-ruut-test | 0,29080014 | | | |

8. Sõnastage lõppjärelendus

(Kas seos on statistiliselt oluline? Miks te nii otsustasite? Milles see seos ikka seisneb?).

Kuigi paremate matemaatika hinnetega tudengite hulgas on autoomanikke rohkem – matemaatikas viielistest tudengitest omavad auto kasutamise võimalust 50,0%, matemaatikas neljalistest 41,7% ja matemaatikas kolmelistest ainult 14,3%, ei ole seos auto omamise ja matemaatika hinde vahel statistiliselt oluline ($p = 0,291 > 0,05$). Seega võime küll rääkida kirjeldatud seosest 2011. aasta veterinaarmeditsiini esimese kursuse tudengite hulgas, aga ei saa seda seost üldistada kõigile Maaülikooli 2011. aasta esimese kursuse tudengitele, selleks on andmeid lihtsalt liiga vähe (eriti vähe on matemaatikas kolmelisi ☺).

Lisa. Statistikapaketid annavad hoiatuse, et χ^2 -testi tulemused ei pruugi olla päris täpsed, kui teoreetiliste (nullhüpoteesile vastavate) sageduste tabelis on üle 20% lahtrite väärtused viiest väiksemad. Antud ülesande puhul on viiest väiksemad poolte lahtrite väärtused. Sestap on teie kursuse andmestik nii või teisiti pisut liiga väike tudengite matemaatika hinde ja auto omamise vahelise seose testimiseks χ^2 -testiga (aga tulemus muideks, on suhteliselt täpne – kasutades ligikaudselt arvatava χ^2 -testi asemel märksa keerukamat, aga sõltumata andmete hulgast täpse p-väärtuse välja rehkendavat Fisheri täpset testi, on tulemuseks $p = 0,303$, mis ei erine kuigivõrd χ^2 -testi ligikaudselt tulemusest $p = 0,291$).