

## 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 sugu ja auto omamine on seotud, st kas autot omavate tudengite arv on mees- ja naistudengite hulgas erinev?**

1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'SUGU' 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 kahemõõtmeline sagedustabel tunnustele 'SUGU' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi;
  6. arvutage viimase alusel nullhüpoteesile vastavad (seost ei ole) 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 'SUGU' ja 'AUTO'.

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

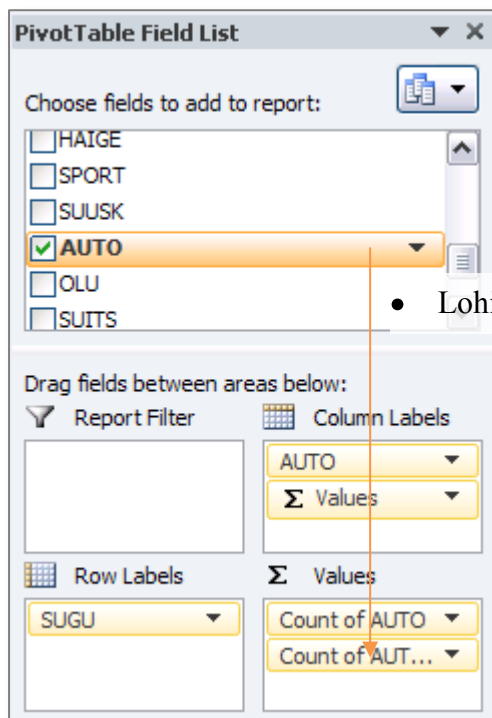
Loodav tabel paigutage samale 'Praks7'-lehele:

The image shows a sequence of steps in Microsoft Excel to create a PivotTable. 
 1. The 'Insert' tab is active, and the 'PivotTable' button is highlighted. 
 2. The 'Create PivotTable' dialog box is open, showing the data source as 'Praks7!\$A\$1:\$5\$53' and the location as 'Praks7!\$V\$1'. 
 3. The 'PivotTable Field List' task pane is shown with 'SUGU' and 'AUTO' selected. 
 4. A red text box notes: 'Näiteks. Samas võib lasta Excelil kokku lugeda ka mõne teise kõigil tudengitel registreeritud tunnuse väärtused.' 
 5. The final PivotTable layout is shown with 'SUGU' as a Row Label and 'AUTO' as a Column Label, with the value field set to 'Count of AUTO'.

Tulemus:

| Count of AUTO      | Column Lab |           |             |
|--------------------|------------|-----------|-------------|
| Row Labels         | ei         | jah       | Grand Total |
| M                  | 6          | 10        | 16          |
| N                  | 18         | 18        | 36          |
| <b>Grand Total</b> | <b>24</b>  | <b>28</b> | <b>52</b>   |

- Leidke tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii realkui ka veeruprotsendid).

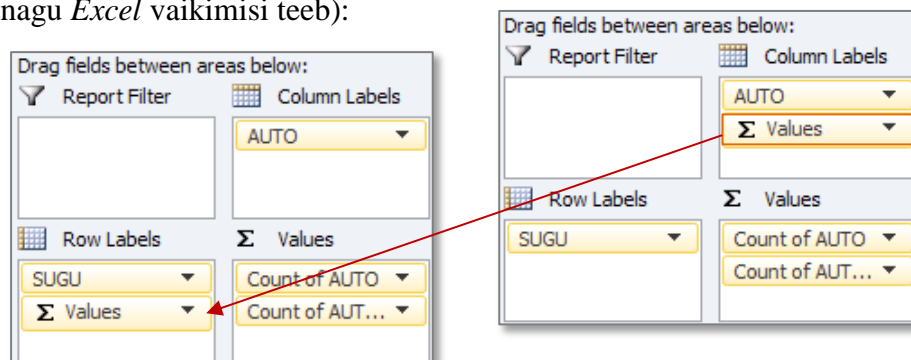


- Lohistage tunnus 'AUTO' ka teine kord väärtuste lahtrisse.

Tulemus:

| Column Labels      | ei            |                | jah           |                | Total Count of AUT | Total Count of AUT |
|--------------------|---------------|----------------|---------------|----------------|--------------------|--------------------|
| Row Labels         | Count of AUTO | Count of AUTO2 | Count of AUTO | Count of AUTO2 |                    |                    |
| M                  | 6             | 6              | 10            | 10             | 16                 | 16                 |
| N                  | 18            | 18             | 18            | 18             | 36                 | 36                 |
| <b>Grand Total</b> | <b>24</b>     | <b>24</b>      | <b>28</b>     | <b>28</b>      | <b>52</b>          | <b>52</b>          |

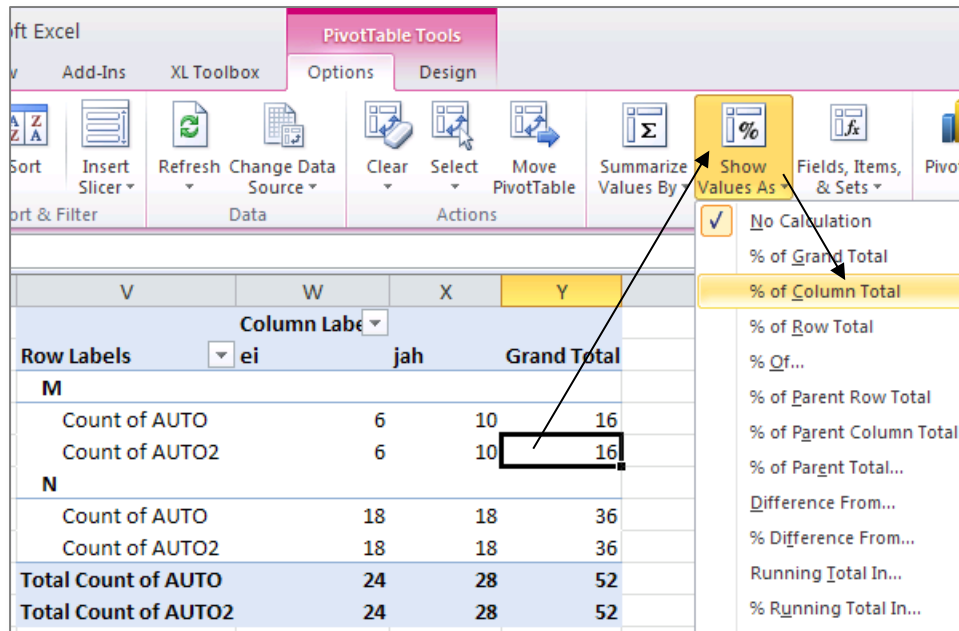
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu Excel vaikimisi teeb):



Tulemus:

| Row Labels                  | ei        | jah       | Grand Total |
|-----------------------------|-----------|-----------|-------------|
| <b>M</b>                    |           |           |             |
| Count of AUTO               | 6         | 10        | 16          |
| Count of AUTO2              | 6         | 10        | 16          |
| <b>N</b>                    |           |           |             |
| Count of AUTO               | 18        | 18        | 36          |
| Count of AUTO2              | 18        | 18        | 36          |
| <b>Total Count of AUTO</b>  | <b>24</b> | <b>28</b> | <b>52</b>   |
| <b>Total Count of AUTO2</b> | <b>24</b> | <b>28</b> | <b>52</b>   |

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



Tulemus:

| Row Labels                  | ei        | jah       | Grand Total |
|-----------------------------|-----------|-----------|-------------|
| <b>M</b>                    |           |           |             |
| Count of AUTO               | 6         | 10        | 16          |
| Count of AUTO2              | 6         | 10        | 16          |
| <b>N</b>                    |           |           |             |
| Count of AUTO               | 18        | 18        | 36          |
| Count of AUTO2              | 18        | 18        | 36          |
| <b>Total Count of AUTO</b>  | <b>24</b> | <b>28</b> | <b>52</b>   |
| <b>Total Count of AUTO2</b> | <b>24</b> | <b>28</b> | <b>52</b>   |

- Analoogselt arvutage ka reaprotsendid ... ( % of Row Total ).  
(alustage sarnaselt punktis 2 kirjeldatule)

| Row Labels                  | ei             | jah            | Grand Total    |
|-----------------------------|----------------|----------------|----------------|
| <b>M</b>                    |                |                |                |
| Count of AUTO               | 6              | 10             | 16             |
| Count of AUTO2              | 25.00%         | 35.71%         | 30.77%         |
| Count of AUTO3              | 37.50%         | 62.50%         | 100.00%        |
| <b>N</b>                    |                |                |                |
| Count of AUTO               | 18             | 18             | 36             |
| Count of AUTO2              | 75.00%         | 64.29%         | 69.23%         |
| Count of AUTO3              | 50.00%         | 50.00%         | 100.00%        |
| <b>Total Count of AUTO</b>  | <b>24</b>      | <b>28</b>      | <b>52</b>      |
| <b>Total Count of AUTO2</b> | <b>100.00%</b> | <b>100.00%</b> | <b>100.00%</b> |
| <b>Total Count of AUTO3</b> | <b>46.15%</b>  | <b>53.85%</b>  | <b>100.00%</b> |

**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).

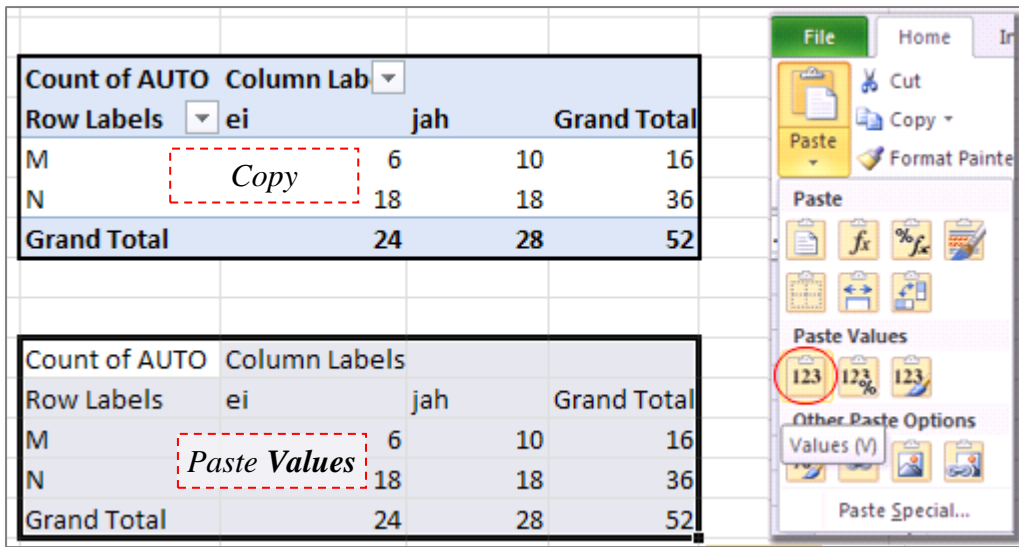
5. Tehke uus 2-mõõtmeline sagedustabel tunnustele 'SUGU' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi.

| Column Lab   |                |                |                |
|--|----------------|----------------|----------------|
| Row Labels   | ei             | jah            | Grand Total    |
| <b>M</b>   |                |                |                |
| Count of AL  | 6              | 10             | 16             |
| Count of AL  | 25.00%         | 35.71%         | 30.77%         |
| Count of AL  | 37.50%         | 62.50%         | 100.00%        |
| <b>N</b>   |                |                |                |
| Count of AL  | 18             | 18             | 36             |
| Count of AL  | 75.00%         | 64.29%         | 69.23%         |
| Count of AL  | 50.00%         | 50.00%         | 100.00%        |
| <b>Total Count of A</b>                              | <b>24</b>      | <b>28</b>      | <b>52</b>      |
| <b>Total Count of A</b>                              | <b>100.00%</b> | <b>100.00%</b> | <b>100.00%</b> |
| <b>Total Count of A</b>                              | <b>46.15%</b>  | <b>53.85%</b>  | <b>100.00%</b> |
| Kommentaarid. ....                                   |                |                |                |
| H <sub>0</sub> : Sugu ja auto omamine ei ole seotud. |                |                |                |
| H <sub>1</sub> : Sugu ja auto omamine on seotud.     |                |                |                |
| <b>Count of AUTO</b> Column Lab                      |                |                |                |
| Row Labels   | ei             | jah            | Grand Total    |
| M  | 6              | 10             | 16             |
| N  | 18             | 18             | 36             |
| <b>Grand Total</b>                                   | <b>24</b>      | <b>28</b>      | <b>52</b>      |

- 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:



- seejärel kustutage ära kopeeritud tabeli **sisu**

| Count of AUTO | Column Labels | ei | jah | Grand Total |
|---------------|---------------|----|-----|-------------|
| M             |               | 6  | 10  | 16          |
| N             |               | 18 | 18  | 36          |
| Grand Total   |               | 24 | 28  | 52          |

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

| Count of AUTO | Column Labels | ei             | jah | Grand Total |
|---------------|---------------|----------------|-----|-------------|
| M             |               | $=Y31*W33/Y33$ |     | 16          |
| N             |               |                |     | 36          |
| Grand Total   |               | 24             | 28  | 52          |

$n_{.1}$  (green),  $n_{1.}$  (blue),  $n$  (purple)

| Count of AUTO | Column Labels | ei          | jah            | Grand Total |
|---------------|---------------|-------------|----------------|-------------|
| M             |               | 7.384615385 | 8.615384615    | 16          |
| N             |               | 16.61538462 | $=Y32*X33/Y33$ | 36          |
| Grand Total   |               | 24          | 28             | 52          |

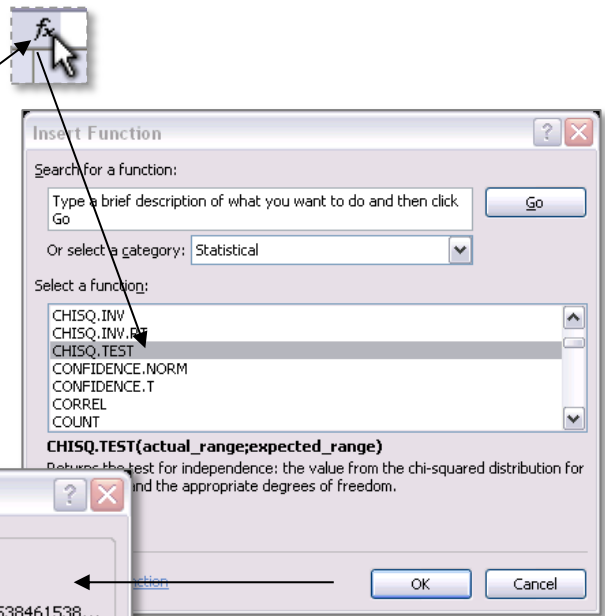
$n_{.2}$  (green),  $n_{2.}$  (blue),  $n$  (purple)

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

Et oleks lihtsam aru saada, milliste arvude võrdlemisel  $\chi^2$ -test baseerub (ehk siis millised tabelite osad tuleb Excel'i funktsioonile CHISQ.TEST ette anda), võib vastavad lahtrid selguse mõttes ära värvida ja ka tabelitele peale kirjutada, mida neis olevad arvud näitavad.

|    | V  | W           | X           | Y           |
|----|--|-------------|-------------|-------------|
| 21 | <u>Tegelikud andmete alusel arvatud (empiirilised) sagedused</u> |             |             |             |
| 22 | Count of AUTO Column Lab   |             |             |             |
| 23 | Row Labels   | ei          | jah         | Grand Total |
| 24 | M  | 6           | 10          | 16          |
| 25 | N  | 18          | 18          | 36          |
| 26 | Grand Total  | 24          | 28          | 52          |
| 27 | <u>Teoreetilised nullhüpooteesile vastavad sagedused</u>         |             |             |             |
| 29 | Count of AUTO Column Labels                                      |             |             |             |
| 30 | Row Labels   | ei          | jah         | Grand Total |
| 31 | M  | 7.384615385 | 8.615384615 | 16          |
| 32 | N  | 16.61538462 | 19.38461538 | 36          |
| 33 | Grand Total  | 24          | 28          | 52          |
| 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).



**Function Arguments**

CHISQ.TEST

**Actual\_range** W24:X25 = {6;10;18;18}

**Expected\_range** W31:X32 = {7.38461538461539;8.61538461538...}

= 0.403986354

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

**Actual\_range** is the range of data that contains observations to test against expected values.

Formula result = 0.403986354

[Help on this function](#)

NB! Vanemais Exceli versioonides testib sama hüpooteesi funktsioon CHITEST.

Tulemus:

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

Kui noormeeste hulgas on autoomanikke 62,5%, siis neidude hulgas vaid 50,0%. Samas, kõigist autoomanikest on 35,7% mehed ja 64,3% naised (neidusid on kursusel lihtsalt rohkem). Hii-ruut-testi kohaselt ei ole kirjeldatud erinevus siiski statistiliselt oluline ( $p = 0,404 > 0,05$ ), st et hoolimata suuremast autoomanike hulgast noormeeste hulgas, ei saa antud andmestiku alusel lugeda tudengi soo ja auto olemise seotust tõestatuks (väites seos olemasolu, eksiksime 40,4%-lise tõenäosusega, mis on enam, kui traditsiooniliselt lubatav 5% eksimist; ehk teisisõnu – andmetest ilmnenud seos auto olemise ja soo vahel võib 40,4%-lise tõenäosusega olla tekkinud ka lihtsalt juhuslikult).

|  |  |             |             |             |
|--|--|-------------|-------------|-------------|
|  |  |             |             |             |
|  | H <sub>0</sub> : Sugu ja auto omamine ei ole seotud.             |             |             |             |
|  | H <sub>1</sub> : Sugu ja auto omamine on seotud.                 |             |             |             |
|  | <u>Tegelikud andmete alusel arvatud (empiirilised) sagedused</u> |             |             |             |
|  | Count of AUTO Column Lab   |             |             |             |
|  | Row Labels   | ei          | jah         | Grand Total |
|  | M  | 6           | 10          | 16          |
|  | N  | 18          | 18          | 36          |
|  | Grand Total  | 24          | 28          | 52          |
|  | <u>Teoreetilised nullhüpooteesile vastavad sagedused</u>         |             |             |             |
|  | Count of AUTO Column Labels                                      |             |             |             |
|  | Row Labels   | ei          | jah         | Grand Total |
|  | M  | 7.384615385 | 8.615384615 | 16          |
|  | N  | 16.61538462 | 19.38461538 | 36          |
|  | Grand Total  | 24          | 28          | 52          |
|  | Hii-ruut-test  | 0.403986354 |             |             |