

## 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 lõppjärelus (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:

**Create PivotTable**

Choose the data that you want to analyze

Select a table or range  
Table/Range: Praks7!\$A\$1:\$5\$54

Use an external data source  
Choose Connection...

Connection name:

Choose where you want the PivotTable report to be placed

New Worksheet  
 Existing Worksheet  
Location: Praks7!\$V\$1

OK Cancel

**PivotTable Field List**

Choose fields to add to report:

RIIK  
 SUGU  
 PIKKUS  
 MASS  
 PEA\_P  
 JALANR  
 ODE\_VEND  
 MAT\_HINNE  
 HOMMIK  
 PUDER  
 LEMMIK  
 HAIGE  
 SPORT  
 SUUSK  
 AUTO

Drag fields between areas below:

Report Filter  
Column Labels  
Row Labels  
Values

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

**Drag fields between areas below:**

Report Filter  
Column Labels: AUTO  
Row Labels: SUGU  
Values: Count of AUTO

Tulemus:

Count of AUTO	Column Labels		
Row Labels	ei	jah	Grand Total
N	21	23	44
M	1	7	8
(blank)	1		1
<b>Grand Total</b>	<b>23</b>	<b>30</b>	<b>53</b>

- Kaotage tabelist puuduvale väärtustele (blank) vastav rida (tudeng, kelle sugu on teadmata).

Count of AUTO		Column Labels		
Row Labels	ei	jah	Grand Total	
	21	23	44	
	1	7	8	
	1		1	
	23	30	53	

Tulemus:

Count of AUTO		Column Labels		
Row Labels	ei	jah	Grand Total	
N	21	23	44	
M	1	7	8	
<b>Grand Total</b>	<b>22</b>	<b>30</b>	<b>52</b>	

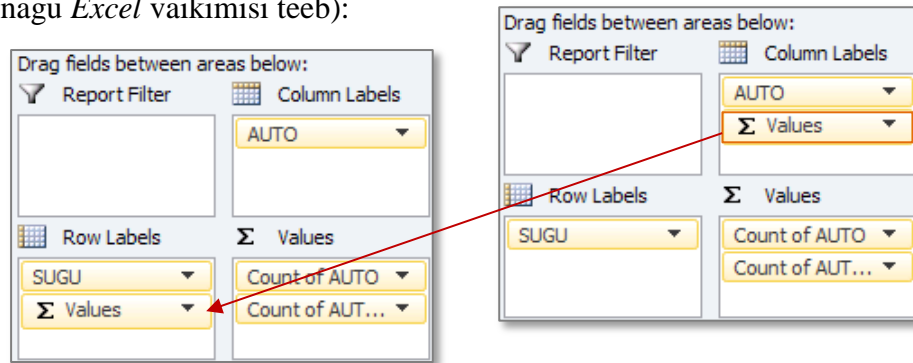
2. Leidke tabelisse lisaks absoluutsetele sagedustele ka kahel suhtelised sagedused (nii reaku kui ka veeruprotsendid).

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

Tulemus:

Column Labels						
ei	jah	Total Count of AU		Total Count of AUT		
Row Labels	Count of AUTO	Count of AUTO2	Count of AUTO	Count of AUTO2		
N	21	21	23	23	44	44
M	1	1	7	7	8	8
<b>Grand Total</b>	<b>22</b>	<b>22</b>	<b>30</b>	<b>30</b>	<b>52</b>	<b>52</b>

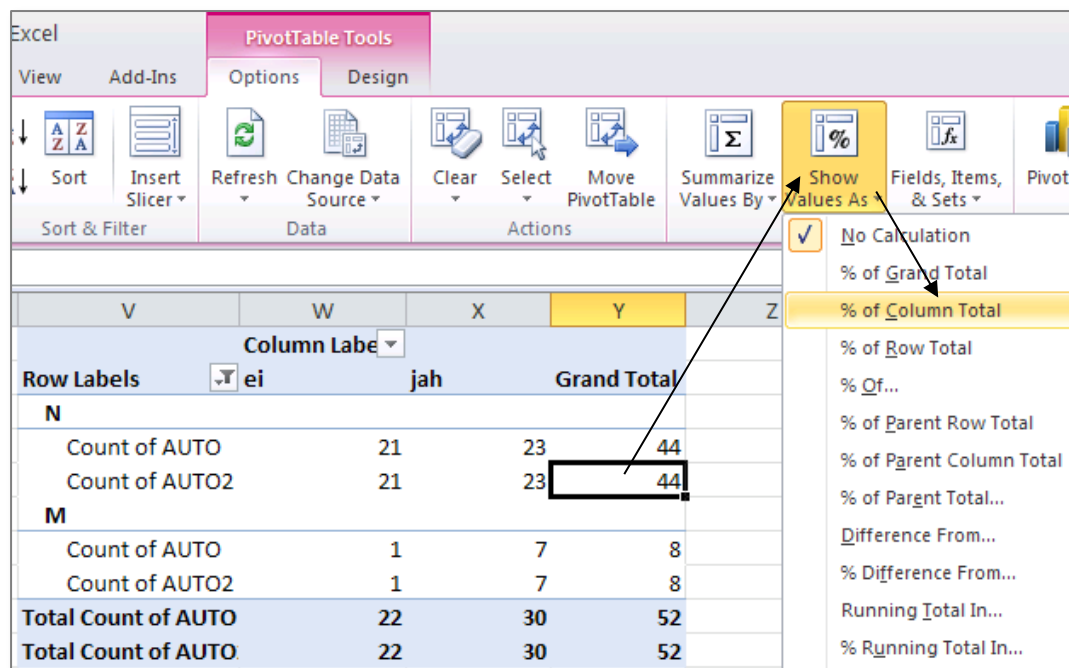
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu *Excel* vaikimisi teeb):



Tulemus:

Row Labels	Column Labels	jah	Grand Total	
<b>N</b>				
Count of AUTO		21	23	44
Count of AUTO2		21	23	44
<b>M</b>				
Count of AUTO		1	7	8
Count of AUTO2		1	7	8
<b>Total Count of AUTO</b>		<b>22</b>	<b>30</b>	<b>52</b>
<b>Total Count of AUTO2</b>		<b>22</b>	<b>30</b>	<b>52</b>

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



Tulemus:

Row Labels	Column Labels	jah	Grand Total	
<b>N</b>				
Count of AUTO		21	23	44
Count of AUTO2		95,45%	76,67%	84,62%
<b>M</b>				
Count of AUTO		1	7	8
Count of AUTO2		4,55%	23,33%	15,38%
<b>Total Count of AUTO</b>		<b>22</b>	<b>30</b>	<b>52</b>
<b>Total Count of AUTO2</b>		<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

- Analoogselt arvutage ka reaprotsendid ... (% of Row Total).

(alustage sarnaselt punktis 2 kirjeldatule)

Row Labels	ei	jah	Grand Total
<b>N</b>			
Count of AUTO	21	23	44
Count of AUTO2	95,45%	76,67%	84,62%
Count of AUTO3	47,73%	52,27%	100,00%
<b>M</b>			
Count of AUTO	1	7	8
Count of AUTO2	4,55%	23,33%	15,38%
Count of AUTO3	12,50%	87,50%	100,00%
<b>Total Count of AUTO</b>	<b>22</b>	<b>30</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>42,31%</b>	<b>57,69%</b>	<b>100,00%</b>

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

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

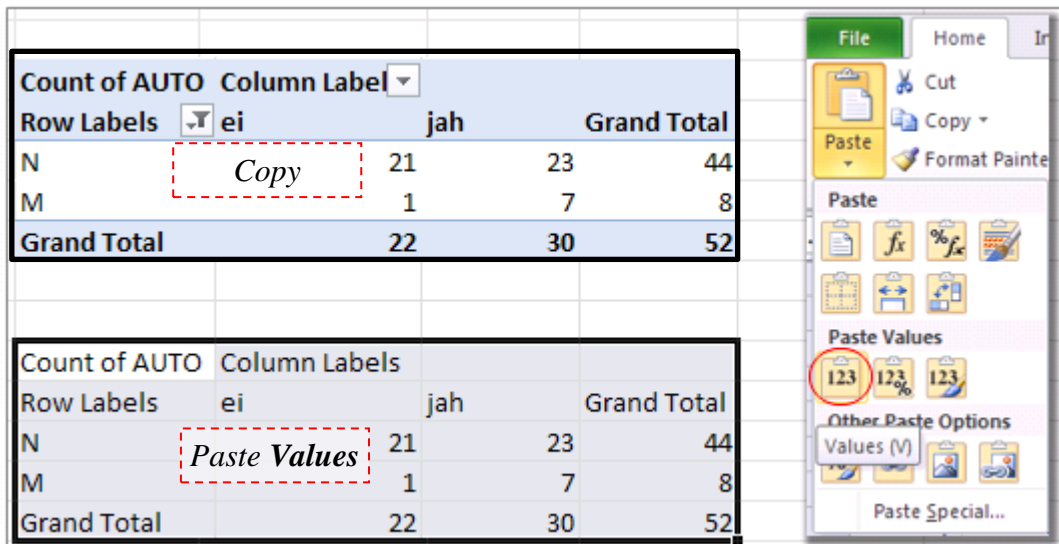
- Tehke uus 2-mõõtmeline sagedustabel tunnustele 'SUGU' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi.

Row Labels	ei	jah	Grand Total
<b>N</b>			
Count of AL	21	23	44
Count of AL	95,45%	76,67%	84,62%
Count of AL	47,73%	52,27%	100,00%
<b>M</b>			
Count of AL	1	7	8
Count of AL	4,55%	23,33%	15,38%
Count of AL	12,50%	87,50%	100,00%
<b>Total Count of A</b>	<b>22</b>	<b>30</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>42,31%</b>	<b>57,69%</b>	<b>100,00%</b>
Kommentaariid. ....			
H <sub>0</sub> : Sugu ja auto omamine ei ole seotud.			
H <sub>1</sub> : Sugu ja auto omamine on seotud.			
Count of AUTO			
Row Labels	ei	jah	Grand Total
N	21	23	44
M	1	7	8
<b>Grand Total</b>	<b>22</b>	<b>30</b>	<b>52</b>

- 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			
Row Labels	ei	jah	Grand Total	
N	21	23	44	
M	1	7	8	
Grand Total	22	30	52	

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

Formula bar:  $=Y31*W33/Y33$

Count of AUTO	Column Labels			
Row Labels	ei	jah	Grand Total	
N	$=Y31*W33/Y33$		44	$n_1$
M			8	
Grand Total	22	30	52	$n$

$n_1$  is indicated by a blue bracket under the 'Grand Total' column for row N.  $n$  is indicated by a purple bracket under the 'Grand Total' row.  $n_1$  is also indicated by a green bracket under the 'Grand Total' row for column 'ei'.

.....

Count of AUTO	Column Labels			
Row Labels	ei	jah	Grand Total	
N	18,61538462	25,38461538	44	$n_2$
M	3,384615385	$=Y32*X33/Y33$	8	
Grand Total	22	30	52	$n$

$n_2$  is indicated by a blue bracket under the 'Grand Total' column for row N.  $n$  is indicated by a purple bracket under the 'Grand Total' row.  $n_2$  is also indicated by a green bracket under the 'Grand Total' row for column 'jah'.

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	Tegelikud andmete alusel arvatud (empiirilised) sagedused			
22	Count of AUTO	Column Label		
23	Row Labels	ei	jah	Grand Total
24	N	21	23	44
25	M	1	7	8
26	Grand Total	22	30	52
27	Teoreetilised nullhüpooteesile vastavad sagedused			
29	Count of AUTO	Column Labels		
30	Row Labels	ei	jah	Grand Total
31	N	18,61538462	25,38461538	44
32	M	3,384615385	4,615384615	8
33	Grand Total	22	30	52
35	Hii-ruut-test			
36				

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

NB! Vanemais Exceli versioonides testib sama hüpooteesi funktsioon CHITEST.

8. Sõnastage lõppjärelaus

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

Kui meesterahvaste hulgas on autoomanikke 87,5%, siis naiste hulgas vaid 52,3%. Samas, kõigist autoomanikest on 23,3% mehed ja 76,7% naised (neidusid on kursusel lihtsalt palju rohkem). Hii-ruut-testi kohaselt ei ole kirjeldatud erinevus siiski statistiliselt oluline ( $p = 0,064 > 0,05$ ), st et hoolimata suuremast autoomanike hulgast noormeeste hulgas, ei saa antud andmestiku alusel lugeda tudengi soo ja auto omamise seotust tõestatuks (väites seos olemasolu eksiksime 6,4%-lise tõenäosusega, mis on enam, kui traditsiooniliselt lubatav 5% eksimist).

Tulemus:

H <sub>0</sub> : Sugu ja auto omamine ei ole seotud.		
H <sub>1</sub> : Sugu ja auto omamine on seotud.		
Tegelikud andmete alusel arvatud (empiirilised) sagedused		
Count of AUTO	Column Label	
Row Labels	ei	jah
N	21	23
M	1	7
Grand Total	22	30
Teoreetilised nullhüpooteesile vastavad sagedused		
Count of AUTO	Column Labels	
Row Labels	ei	jah
N	18,61538462	25,38461538
M	3,384615385	4,615384615
Grand Total	22	30
Hii-ruut-test	0,063574183	