

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:

Create PivotTable

Choose the data that you want to analyze

Select a table or range
Table/Range: Praks7!\$A\$1:\$5\$58

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
- PIKKJUS
- 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: []

Report Filter []
Column Labels SUGU
Row Labels AUTO
Values Count of RIIK

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

Tulemus:

Count of RIIK	Column Labels			
Row Labels	M	N	(blank)	Grand Total
ei	13	23		36
jah	13	6	1	20
(blank)	1			1
Grand Total	27	29	1	57

- Kaotage tabelist puuduvatele väärtustele (blank) vastav rida ja ka veerg.

Count of RIIK		Column Labels		
Row Labels	M	N	(blank)	Grand Total
ei	13	23		36
jah	13	6	1	20
(blank)	1			1
Grand Total	27	29	1	57

Tulemus:

Count of RIIK		Column Labels		Grand Total
Row Labels	M	N		
ei	13	23		36
jah	13	6		19
Grand Total	26	29		55

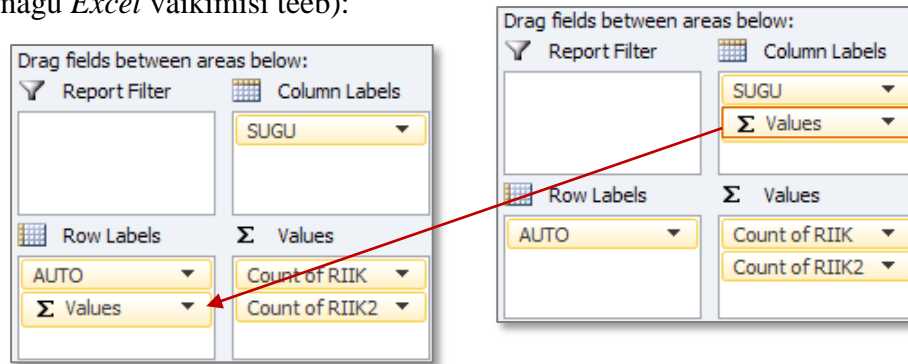
2. Leidke tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii reaku kui ka veeruprotsendid).

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

Tulemus:

Column Labels		Total Count of RIIK				Total Count of RIIK2	
Row Labels	M	N	Count of RIIK	Count of RIIK2	Count of RIIK	Count of RIIK2	
ei	13	13	23	23	36	36	
jah	13	13	6	6	19	19	
Grand Total	26	26	29	29	55	55	

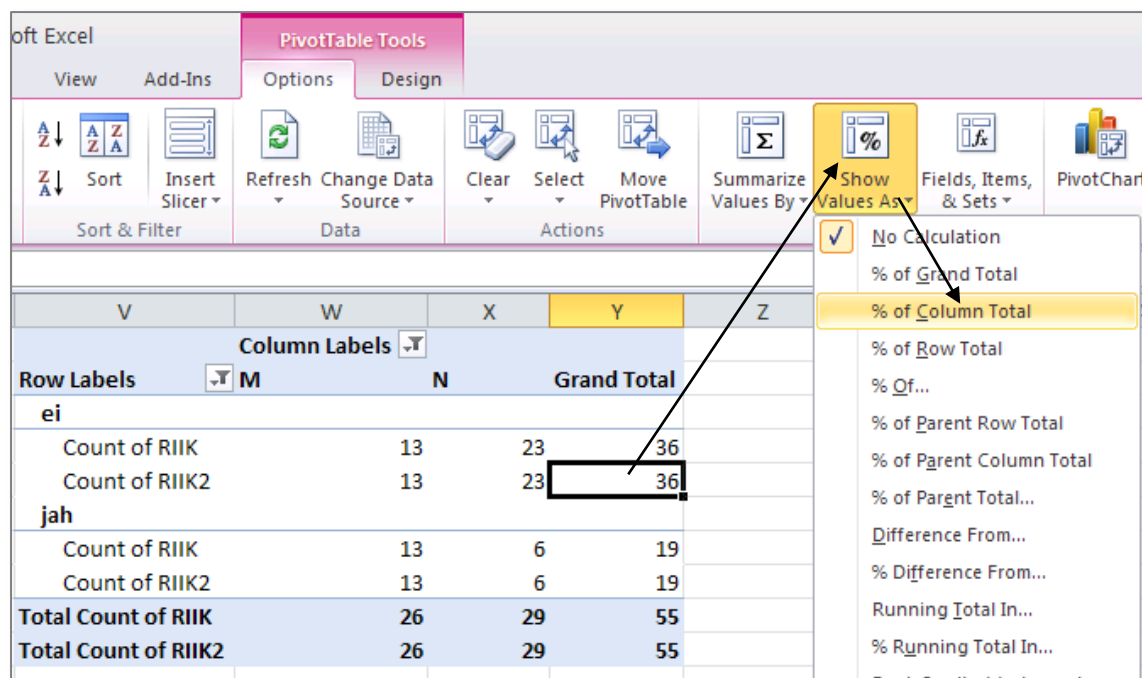
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu *Excel* vaikimisi teeb):



Tulemus:

Row Labels	M	N	Grand Total
ei			
Count of RIIK	13	23	36
Count of RIIK2	13	23	36
jah			
Count of RIIK	13	6	19
Count of RIIK2	13	6	19
Total Count of RIIK	26	29	55
Total Count of RIIK2	26	29	55

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



Tulemus:

Row Labels	M	N	Grand Total
ei			
Count of RIIK	13	23	36
Count of RIIK2	50,00%	79,31%	65,45%
jah			
Count of RIIK	13	6	19
Count of RIIK2	50,00%	20,69%	34,55%
Total Count of RIIK	26	29	55
Total Count of RIIK2	100,00%	100,00%	100,00%

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

Row Labels	M	N	Grand Total
ei			
Count of RIIK	13	23	36
Count of RIIK2	50,00%	79,31%	65,45%
Count of RIIK3	36,11%	63,89%	100,00%
jah			
Count of RIIK	13	6	19
Count of RIIK2	50,00%	20,69%	34,55%
Count of RIIK3	68,42%	31,58%	100,00%
Total Count of RIIK	26	29	55
Total Count of RIIK2	100,00%	100,00%	100,00%
Total Count of RIIK3	47,27%	52,73%	100,00%

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).

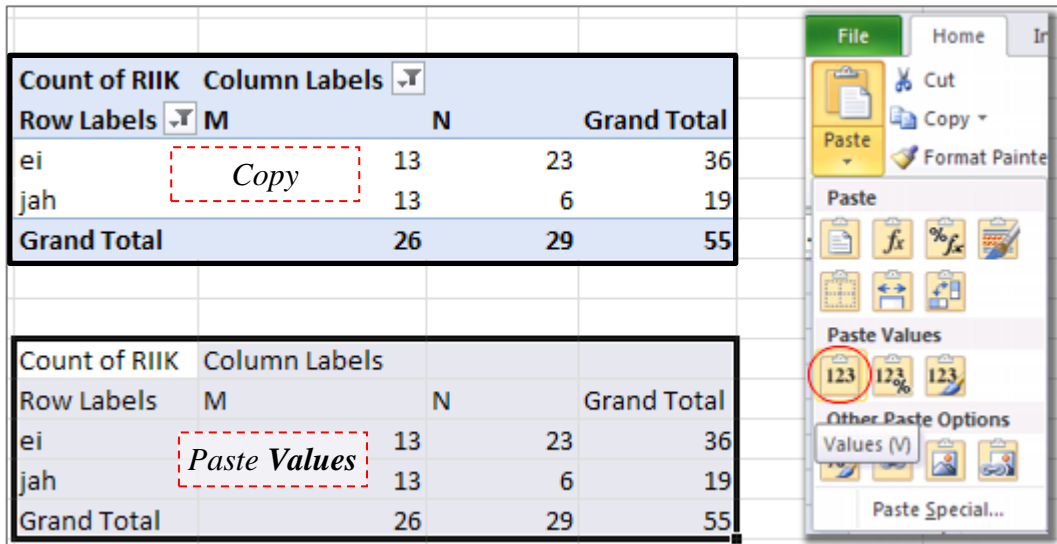
- Tehke uus 2-mõõtmeline sagedustabel tunnustele 'SUGU' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi.

Row Labels	M	N	Grand Total
ei			
Count of R	13	23	36
Count of R	50,00%	79,31%	65,45%
Count of R	36,11%	63,89%	100,00%
jah			
Count of R	13	6	19
Count of R	50,00%	20,69%	34,55%
Count of R	68,42%	31,58%	100,00%
Total Count of	26	29	55
Total Count of	100,00%	100,00%	100,00%
Total Count of	47,27%	52,73%	100,00%
Kommentaariid.			
H ₀ : Sugu ja auto omamine ei ole seotud.			
H ₁ : Sugu ja auto omamine on seotud.			
Count of RIIK			
Row Labels	M	N	Grand Total
ei	13	23	36
jah	13	6	19
Grand Total	26	29	55

- 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 RIIK	Column Labels		
Row Labels	M	N	Grand Total
ei	13	23	36
jah	13	6	19
Grand Total	26	29	55

The data cells (13, 23, 13, 6) are crossed out with a red 'X'.

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

Formula bar: $=Y31*W33/Y33$

	V	W	X	Y
29	Count of RIIK	Column Labels		
30	Row Labels	M	N	Grand Total
31	ei	$=Y31*W33/Y33$		36 n_1
32	jah			19
33	Grand Total	26	29	55 n

Diagrammatical annotations: A blue bracket under '36' is labeled n_1 . A purple bracket under '55' is labeled n . A green bracket under '26' is labeled $n_{.1}$. A purple arrow points from the formula cell to the '36' cell.

.....

	V	W	X	Y
29	Count of RIIK	Column Labels		
30	Row Labels	M	N	Grand Total
31	ei	17,01818182	18,98181818	36 n_1
32	jah	8,981818182	$=Y32*X33/Y33$	19 n_2
33	Grand Total	26	29	55 n

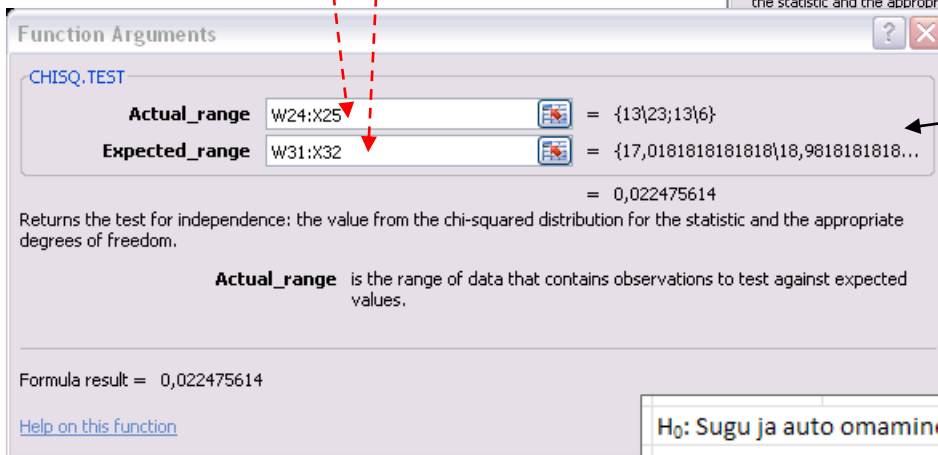
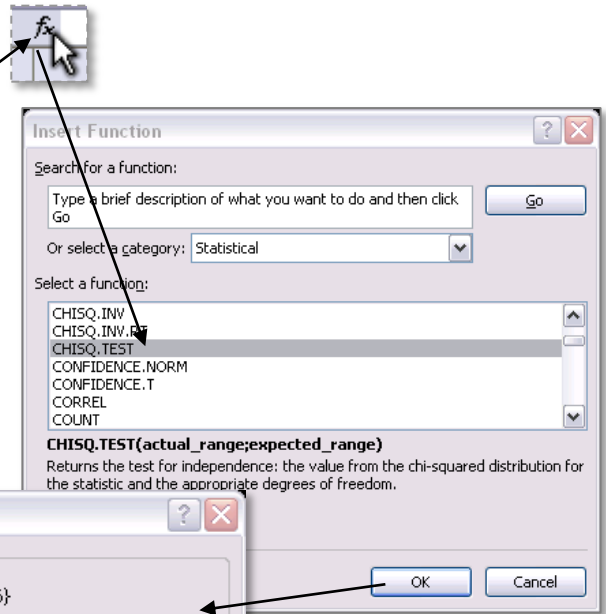
Diagrammatical annotations: A blue bracket under '19' is labeled n_2 . A purple bracket under '55' is labeled n . A green bracket under '29' is labeled $n_{.2}$. A purple arrow points from the formula cell to the '19' cell.

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.

	V	W	X	Y
21				
22	Count of RIIK	Column Lab		
23	Row Labels	M	N	Grand Total
24	ei	13	23	36
25	jah	13	6	19
26	Grand Total	26	29	55
27				
28				
29	Count of RIIK	Column Labels		
30	Row Labels	M	N	Grand Total
31	ei	17,01818182	18,98181818	36
32	jah	8,981818182	10,01818182	19
33	Grand Total	26	29	55
34				
35				
36	Hii-ruut-test			

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 hüpoteesi funktsioon CHITEST.

Tulemus:

8. Sõnastage lõppjärelus

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

Kui meesterahvaste hulgas on autoomanikke 50,0%, siis naiste hulgas vaid 20,7%. Kõigist autoomanikest on 68,4% mehed ja 31,6% naised. Hii-ruut-testi kohaselt on kirjeldatud erinevus ka statistiliselt oluline ($p = 0,022 < 0,05$), st et käsitledes teie kursust valimina kõigist EMÜ esmakursuslastest, võime lugeda tõestatuks, et tudengi sugu ja auto omamine on seotud (ja seda väites on eksimise tõenäosus vaid 2,2%).

H ₀ : Sugu ja auto omamine ei ole seotud.				
H ₁ : Sugu ja auto omamine on seotud.				
<u>Tegelikud andmete alusel arvatud (empiirilised) sagedused</u>				
Count of RIIK	Column Lab			
Row Labels	M	N		Grand Total
ei	13	23		36
jah	13	6		19
Grand Total	26	29		55
<u>Teoreetilised nullhüpoteesile vastavad sagedused</u>				
Count of RIIK	Column Labels			
Row Labels	M	N		Grand Total
ei	17,01818182	18,98181818		36
jah	8,981818182	10,01818182		19
Grand Total	26	29		55
Hii-ruut-test	0,022475614			